

DOCUMENT RESUME

ED 159 195

TM 007 436

AUTHOR Massad, Carolyn Emrick, Ed.
 TITLE Resource Notebook of Information for Assessment and Evaluation: Processes for T & E. No. 6 in a Series of Handbooks on Comprehensive Planning for Local Education Districts. The Evaluation Improvement Program.
 INSTITUTION Educational Testing Service, Princeton, N.J.
 SPONS AGENCY New Jersey State Dept. of Education, Trenton.
 PUB DATE Dec 77
 NOTE 335p.
 EDRS PRICE MF-\$0.83 Plus Postage. HC Not Available from EDRS.
 DESCRIPTORS Achievement Tests; Educational Assessment; *Educational Objectives; Efficiency; Elementary Secondary Education; *Information Sources; Needs Assessment; Program Development; *Program Evaluation; *Program Improvement; Program Planning; School Districts; *State Standards; Test Reviews; *Test Selection
 IDENTIFIERS New Jersey

ABSTRACT

This Resource Notebook is designed to help New Jersey school districts implement the state's Public School Education Act of 1975 which mandates a thorough and efficient (T & E) system of free public schools. The notebook is divided into three parts. Part I, the user's guide, discusses the purpose of data-based decision making, specifically what information is needed and who is involved. Part II focuses on five of the six steps of the T & E continuing assessment-planning-evaluation cycle. These five steps and their related chapters are (1) goal setting; (2) setting objectives and standards; (3) needs assessment; (4) program improvement/program selection; and (5) program evaluation. Part III, appendices, consists of selected references for assessment and evaluation. In appendix A, separate matrices match 243 measurement instruments to the 12 New Jersey Educational Outcome Goals and to variables of test administration and scoring. Informative non-evaluative descriptions of 175 tests that are relevant to one or more of the goals are also included. Appendix B contains statistical and sampling guidelines. Appendix C is a bibliography on test review sources and on educational measurement and evaluation. Concluding the notebook is a glossary and a directory of state and national sources of information and assistance in assessment and evaluation. (CP)

 * Reproductions supplied by EDRS are the best that can be made *
 * from the original document. *

FD159105

U.S. DEPARTMENT OF HEALTH
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY.

William A. ...
...

INFORMATION FOR ASSESSMENT AND EVALUATION



... Evaluation Improvement Program

EDUCATIONAL TESTING SERVICE
PRINCETON, NEW JERSEY



**RESOURCE
NOTEBOOK of
INFORMATION for
ASSESSMENT and
EVALUATION:
PROCESSES for T&E**

1977



**STATE of NEW JERSEY
DEPARTMENT of EDUCATION**

STATE BOARD OF EDUCATION

RUTH H. MANCUSO
President

P. PAUL RICCI
Vice President

S. DAVID BRANDT

WILLIAM COLON

ANNE S. DILLMAN

BRYANT GEORGE

E. CONSTANCE MONTGOMERY

KATHERINE K. NEUBERGER
Chairperson, State Board of Higher Education

SONIA B. RUBY

JACK SLATER

SUSAN N. WILSON

HELEN A. ZEHNER

GLASSBORO

VINELAND

CHERRY HILL

JERSEY CITY

PERTH AMBOY

TEANECK

MORRISTOWN

LINCROFT

ATLANTIC CITY

PATERSON

PRINCETON

WOODSTOWN

Secretary
FRED G. BURKE
Commissioner of Education

Ex Officio
T. Edward Hollander
Chancellor of Higher Education

RESOURCE NOTEBOOK OF INFORMATION FOR ASSESSMENT AND EVALUATION
PROCESSES FOR T & E

Edited by

Carolyn Emrick Massad

Educational Testing Service



The Evaluation Improvement Program

"The materials presented herein were prepared pursuant to a grant from the New Jersey State Department of Education under provisions of Title IV-C of the Elementary and Secondary Education Act (1965), as amended. However, the opinions expressed herein do not necessarily reflect the position or policy of the New Jersey State Department of Education or the United States Office of Education, 1977.

Copyright © 1977 by the Freehold Regional High School District. All rights reserved.

First Edition, December 1977

Published by Educational Testing Service, Princeton, New Jersey



Educational Testing Service, a nonprofit organization, is
an Equal Opportunity Employer

RESOURCE NOTEBOOK OF INFORMATION FOR ASSESSMENT AND EVALUATION
PROCESSES FOR T & E

No. 6 in a Series of Handbooks on Comprehensive Planning
for Local Education Districts

New Jersey State Department of Education

Fred G. Burke
Commissioner of Education

Ralph H. Lataille
Deputy Commissioner

Gary M. Gappert
Assistant Commissioner
Division of Research,
Planning and Evaluation

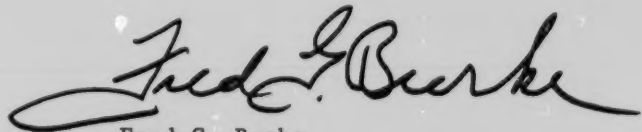
Bernard J. Weiss
Assistant Commissioner
Division of School Programs

FOREWORD

This Resource Notebook of Information for Assessment and Evaluation: Processes for T & E provides a valuable new source of support for educators and other citizens who are engaged in carrying out the provisions of the Public School Education Act of 1975. It is precisely what its title suggests: a resource document that contains guidelines to aid educators and others with some of the more difficult problems and techniques involved in setting goals, objectives, and standards, identifying needs, selecting new programs and improving old ones, and evaluating the results.

Certainly none of the work of providing a thorough and efficient system of public schools in New Jersey can succeed without scrupulous planning based on intelligent decisions. This Resource Notebook will facilitate such planning.

I recommend this notebook to all who are engaged in this important task.

A handwritten signature in dark ink, reading "Fred G. Burke". The signature is fluid and cursive, with a large, sweeping initial "F" and "B".

Fred G. Burke

Commissioner

ACKNOWLEDGEMENT I

The New Jersey Department of Education contracted with Educational Testing Service of Princeton, New Jersey, to develop a resource notebook to help school districts in their planning for "T & E" (the popular way of referring to a "thorough and efficient" process of education). Several individuals at Educational Testing Service have contributed to the preparation of this notebook. Pamela Rosen and her staff of the Test Collection were primarily responsible for the test descriptions, matrices of test measures, and the selected bibliography of sources for test reviews. Michael Zieky is the author of Chapter 2, "Objective Setting and Standard Setting" and the appendix on selecting a sample and interpreting assessment results for the sample. Nathaniel Hartshorne wrote Chapter 4, "Program Improvement/Program Selection," and edited the final copy of the User's Guide and Part II. The major portion of Chapter 5 was derived from material prepared by Dr. Edmond Weiss while he was employed at the Educational Improvement Center - South Jersey Region. Estelle Bartels edited Part III. The project director commented on and edited drafts of the above contributions and prepared the remaining portions of the notebook with assistance from Eleanor Clemson, Wonnice Leonie, David Reggi, and Janet Zahn.

The New Jersey Department of Education was continually involved in the planning and reviewing of materials. Special acknowledgment is given to Evelyn Ogden, who had primary responsibility for the project, and to Charles Barthe, Irv Cantor, Thomas Corcoran, Jean Sadenwater, Art Spangenburg, and Dottie Wilson for their contributions.

Assistance from others who served as a panel of reviewers for the plans and draft copy of the notebook is also gratefully acknowledged. The panel included: David Adler, June Coultas, Virgil Engels, Patricia Horton, Frank Johnson, Ronald Leshner, David Pearsall, Thomas Rookey, Gustav Ruh, Harold Seltzer, Willa Spicer, Glenn Tecker, and Bruce Tuckman.

Carolyn Emrick Massad
Project Director

ACKNOWLEDGEMENT II

Rapid publication and distribution of the Resource Notebook has been made possible through quick and decisive action on the part of Educational Testing Service in response to expressed needs for these materials in the districts and the schools of New Jersey. Jack R. Childress, James R. Deneen, Carolyn E. Massad, and Marc E. Minardi helped expedite the review process within ETS.

The New Jersey Department of Education will have worked closely with ETS to assure that known needs are met as these materials find their way into wide use. As needs for in-service workshops in this area become apparent, for example, details will be worked out to make it possible for T & E process workshops to be scheduled. Evelyn Ogden and William Volk have been most helpful in pointing our thinking in such directions.

The active involvement on the part of Educational Improvement Center-Central Jersey Region and of its Materials Distribution Center, in the dissemination-diffusion phase of this effort is gratefully acknowledged. Thomas Rookey, the Executive Director, has been most helpful in letting the potential readership for this volume know of its existence. We also thank the Freehold Regional High School District, Dr. Victor Crespy, Superintendent, for the foresight in seeing a way of at once filling local needs and comparable needs among sister school districts in the State of New Jersey.

William H. Bronson and Alexander I. Law of the California Department of Education look with favor on our broadening the scope of the Evaluation Improvement Program to carry such publications as this one, along with those developed in California as part of its Evaluation Improvement Project. This may well herald the beginning of a concerted effort to advance the state of the art of program evaluation among educational practitioners from coast to coast.

Wesley W. Walton, Director
Evaluation Improvement Program

CONTENTS

FOREWORD BY FRED G. BURKE	iv
ACKNOWLEDGEMENT I	v
ACKNOWLEDGEMENT II	vi
I. USER'S GUIDE	1
II. ORGANIZING FOR EDUCATIONAL IMPROVEMENT	9
Chapter	
1. Goal Setting	11
2. Objective Setting and Standard Setting	23
3. Needs Assessment	41
4. Program Improvement/Program Selection	69
5. Program Evaluation	83
III. APPENDICES	97
A-1. Matrix of Instruments Matched to New Jersey State Goals	99
A-2. Matrix of Test Descriptions	113
A-3. Compendium of Descriptions of Tests (External Measures)	145
A-4. Publisher Addresses	229
B-1. Selecting a Sample and Interpreting Assessment Results from the Sample	233
B-2. Some Short-Cut Statistics for Internal Measures . .	239
B-3. Evaluation of Program Effect in Student Achievement	245
C-1. Selected Sources of Test (External Measures) Reviews	247
C-2. Selected References on Educational Measurement and Evaluation	259
D. Glossary of Selected Terms	267
E. Potential Sources of Information and/or Assistance in Assessment and Evaluation	287

INFORMATION FOR ASSESSMENT AND EVALUATION

PART I

USER'S GUIDE

 **The Evaluation Improvement Program**

With the passage of the Public Education Act of 1975, the New Jersey Legislature moved toward insuring for all children in New Jersey, regardless of socioeconomic status or geographic location, the educational opportunity which will prepare them to function politically, economically, and socially in a democratic society. This was the Legislature's response to the New Jersey constitutional mandate to:

Provide for the maintenance and support of
a thorough and efficient system of free
public schools . . .

Recognizing that the thoroughness and efficiency (T & E) of school systems is a condition that will vary with current and changing needs of both the state and individual school districts, the law emphasizes the necessity for democratic decision making in the local district. Basically, there must be involvement by citizens of the community in setting local goals. There must be a process for setting local objectives and standards and for identifying needs to provide a base-line point of reference for the school district. Most important, the law emphasized the need to develop and implement plans for improvement where needs are made apparent.

About This Notebook

The purpose of this Resource Notebook is to discuss various aspects of the decision-making processes related to T & E and to offer selected references. The notebook is divided into three parts. Part I, the User's Guide, describes this book, its function and structure, and discusses the purpose of data-based decision making. It responds to the questions: What information is needed to make decisions? and Who is involved in the decision-making process?

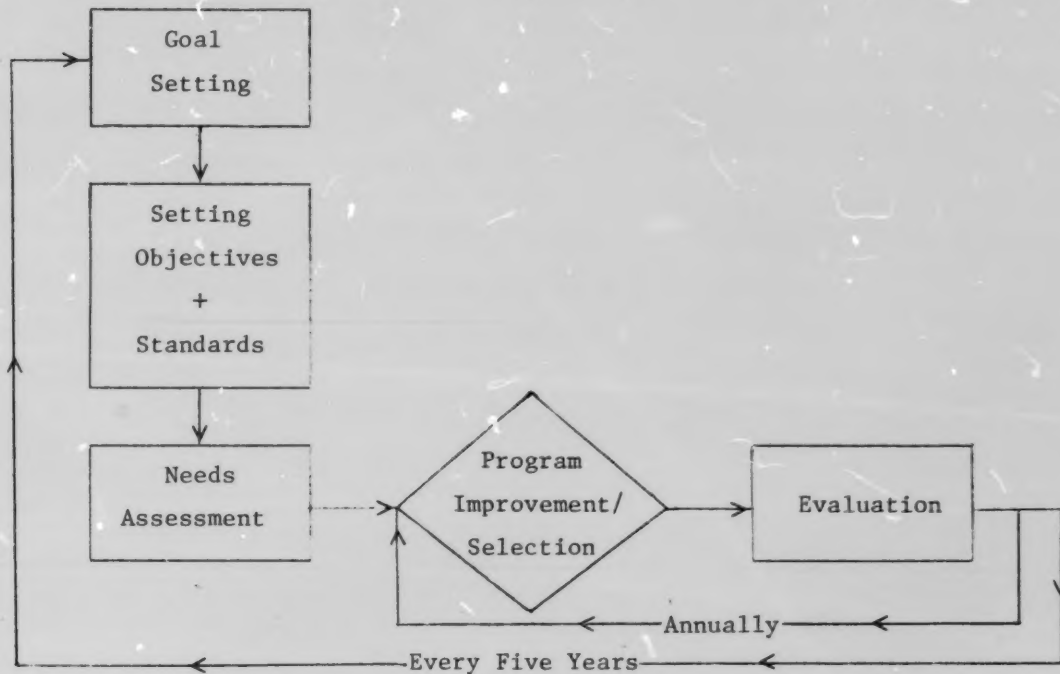
Part II focuses on five of the six steps of the T & E process. These five steps and their related chapters are:

- (1) planning and conducting a goal-setting process with community involvement (Chapter 1)
- (2) planning and conducting a process for setting objectives and standards (Chapter 2)
- (3) planning and conducting a needs-identification process to assess where pupils are at present in relation to the objectives and standards (Chapter 3)
- (4) improving educational programs to achieve objectives (Chapter 4)
- (5) measuring the effectiveness of educational programs in achieving those objectives and making changes as needed (Chapter 5)

Part III consists of selected references including test descriptions, specific references or sources of additional information, and a matrix of measures matched to state outcome goals.

Carrying Out the T & E Process

The five steps of the T & E process consist of a continuing assessment-planning-evaluation cycle that looks like this:



It is expected that in implementing T & E, districts will be at different points in the process at different times. Moreover, within a district, time lines might differ for implementation. For example, objectives for some goals may be in the process of being set while assessment or improvement programs are being carried out for other goal areas. On the next page is an example of a time line a school district might carry out over a year.

Of course, regardless of the time line, for realistic planning the budget must be considered in determining what is feasible for a school district, with careful thought given to both long-range and short-term plans.

SAMPLE OF A SCHOOL DISTRICT'S TIME LINE OVER ONE YEAR

Spring	Fall	Winter	Summer
Set goals.	Do needs assessment for goals.	Set objectives for goals.	Review program-improvement needs for objectives related to goals.
Set objectives and standards for goals.	Set additional objectives for goals.	Assess objectives related to goals.	
Revise remedial program in computational skills at the elementary level.			Evaluate new computational skills program.

What Is Data-Based Decision Making?

Data-based decision making is a process that people use every day, frequently without being aware of it. Every time a judgment is made, even if it is merely arriving at an opinion about something, data-based decision making has taken place: Available information has been considered before the judgment is made.

The use of data-based decision making is most obvious when making major decisions, which, in most instances, are not easily reached. In making a decision, we are confronted with two or more choices: maintaining the status quo or selecting from a number of other options. In making a major decision, a wise person will study the situation carefully--even when one option appears to stand out as the most attractive. Before making such a decision, that person will carefully search for and gather all available

evidence and then form a judgment on which to act. Decisions about education affect the lives of individuals, the community, our nation, and our culture. Therefore, in arriving at any such decisions, it is vitally important to gather as much relevant information as is feasible and practical.

Essentially, data-based decision making in education involves:

(a) clarifying goals and objectives to the point of describing what is wanted and/or needed and at what level of attainment; (b) developing and/or using different ways of getting evidence, or data, about the environment in which the school functions and about the changes in students as a result of school and/or community influences; (c) summarizing and interpreting the data and any other pertinent information; and (d) using that information to make decisions for improving curriculum, teaching, and guidance.

School systems look to the community, their professional staffs, and students both for help in the decision-making process and as resources from whom data for decision making are acquired.

Professional Judgment and Data-Based Decision Making

Once possible kinds of data that are useful in decision making have been explored, the professional educator provides the guidance needed for acquiring and dealing with the data. By virtue of training, experience, and interests, professional educators are needed to help make decisions concerning education. The New Jersey Administrative Code recognizes the importance of the professional educator by stating that:

- (1) The chief school administrator and teaching staff members are to be consulted by the district board of education in the development of a written educational plan for the district (NJAC 6:8-3.1).
- (2) The chief school administrator is to direct the development of district goals in consultation with teaching staff members, pupils, parents, and other district residents (NJAC 6:8-3.2).
- (3) The chief school administrator is to direct the development

of objectives and standards in consultation with teaching staff members (NJAC 6:8-3.3).

- (4) Teaching staff members are to assess student needs to determine attainment of educational objectives (NJAC 6:8-3.4).
- (5) The district educational program (curriculum) is to be developed under the direction of the chief school administrator in consultation with the teaching staff (NJAC 6:8-3.4).
- (6) Teaching staff members are to provide instruction in order to achieve goals, objectives, and standards (NJAC 6:8-3.6).
- (7) A continuous and comprehensive review of pupil progress toward district and school goals and program objectives is to be conducted by teaching staff members under the direction of the chief school administrator, including consultation with parents or guardians and pupils (NJAC 6:8-3.7).

Professional educators are generally in a position to identify or seek out available data, to determine what additional data need to be collected, and to interpret the data acquired for decision making. Moreover, professional educators working with an involved community contribute the leadership necessary for using the data to make wise decisions about setting local goals.

The professional staff can use the data to help school boards set objectives and standards; to identify needs; to plan for program improvement; and to evaluate progress toward achieving goals, objectives, and standards. Further, professional educators implement the decisions made and, therefore, are attuned to their everyday as well as their long-range impact. At every step along the way, the professional judgment of the educator is critical for making and implementing decisions if the educational needs of students and communities are to be met. To provide a "thorough and efficient" educational system requires professional educators organizing for educational improvement and taking an active part in the educational process.

INFORMATION FOR ASSESSMENT AND EVALUATION

PART II

ORGANIZING FOR EDUCATIONAL IMPROVEMENT



The Evaluation Improvement Program

9

25

INFORMATION FOR ASSESSMENT AND EVALUATION

CHAPTER 1

GOAL SETTING

Purpose of This Chapter

The major purpose of this chapter is to discuss goal setting in light of New Jersey's T & E legislation. In the chapter, goal setting is defined, the role of data-based decision making in the goal-setting process is discussed along with kinds and sources of data and the need for pre- and postassessment of this process.

Goal Setting Defined

A goal is a written statement, in general terms, of educational aspirations for learner achievement and the educational process. The primary goal of education in New Jersey is that provided by the Legislature in the Public Education Act of 1975: "to provide to all children in New Jersey, regardless of socioeconomic status or geographical location, the educational opportunity which will prepare them to function politically, economically and socially in a democratic society." As one step toward achieving this goal, procedures were established within the body of the law requiring the State Board of Education with the Commissioner of Education to (a) establish goals and standards which are to be applicable to all public schools and (b) make rules concerning procedures for the establishment of goals, objectives, and standards by local boards of education.

The State Goals. The state goals, determined by the State Board of Education, are defined as outcome and process goals and are applicable to all public schools in the State of New Jersey (NJAC 6:8-2.1). The educational outcome goals are to enable every pupil to:

- (1) Acquire basic skills in obtaining information, solving problems, thinking critically, and communicating effectively

- (2) Acquire a stock of basic information concerning the principles of the physical, biological, and social sciences, the historical record of human achievements and failures and current social issues
- (3) Become an effective and responsible contributor to the decision-making processes of the political and other institutions of the community, state, country, and world
- (4) Acquire the knowledge, skills, and understanding that permit him/her to play a satisfying and responsible role as both producer and consumer
- (5) Acquire job entry-level skills and also acquire knowledge necessary for further education
- (6) Acquire the understanding of, and the ability to form, responsible relations with a wide range of other people, including but not limited to those with social and cultural characteristics different from his/her own
- (7) Acquire the capacities for playing satisfying and responsible roles in family life
- (8) Acquire the knowledge, habits, and attitudes that promote personal and public health, both physical and mental
- (9) Acquire the ability and the desire to express himself/herself creatively in one or more of the arts and to appreciate the aesthetic expressions of other people
- (10) Acquire an understanding of ethical principles and values and the ability to apply them to his/her own life
- (11) Develop an understanding of his/her own worth, abilities, potentialities, and limitations
- (12) Learn to enjoy the process of learning and to acquire the skills necessary for a lifetime of continuous learning and adaptation to change

These goals are directed toward the learners' acquisition of content, skills, or abilities. The process goals are concerned mainly with how the outcome goals are achieved. The process goals state that the public schools shall

provide:

- (1) Instruction which bears a meaningful relationship to the present and future needs and/or interests of pupils
- (2) Significant opportunities, consistent with the age of the pupil, for helping to determine the nature of the educational experiences of the pupil
- (3) Specialized and individualized kinds of educational experiences to meet the needs of each pupil
- (4) Opportunities for teaching staff members and pupils to make recommendations concerning the operation of the schools
- (5) Comprehensive guidance facilities and services for each pupil
- (6) An environment in which any competition among pupils is positive
- (7) Resources for education, used with maximum efficiency
- (8) Teaching staff members of high quality
- (9) Diverse forms of constructive cooperation with parents and community groups

Setting Local Goals. Recognizing the need for local autonomy, the Public Education Act also requires every school district to develop written local educational goals that are based on district educational needs and, at the same time, are consistent with the intent of the state goals. These local goals are to serve as the basis for the educational program or curriculum for each school. Furthermore, the setting of local goals is to be achieved in consultation with the school district's staff, pupils, parents or guardians of pupils, and other district residents under the direction of the chief school administrator who is responsible for preparing a written plan and calendar for the goal-setting process. In other words, local boards of education must set goals and involve the community, professional educators, and pupils in the process so that the goals will meet the needs of the local school district as well as the requirements of the state.

The Role of Data-Based Decision Making in Goal Setting

Without adequate, well-organized information about the community and school system, however, decision makers may set goals that do not meet appropriate local needs, even though the entire community may have been invited to participate in the project. For example, suppose community meetings are held for people to air their opinions about the local educational system. A predominant complaint raised at these meetings is that students currently enrolled in college have not been adequately prepared by the school system. As a result, the decision makers set this goal: to enable students to acquire the skills and knowledge required for college entrance. This goal, however, carries with it the assumption that all students need these skills and knowledge, which may not be the case. Perhaps one-fourth to one-third of the students plan to attend a technical training school or enter the job market upon graduation from high school and may need and/or profit more from skills and knowledge other than what is required for college entrance. This fact was not brought out at the community meeting, but the information was available from records on past graduating classes and/or guidance counselors. However, because all relevant information was not considered at the same time, the goal set did not adequately reflect the desires of the community. A more appropriate goal might have been: to enable students to acquire the skills and knowledge required for entering college and/or technical training school and/or the job market.

Data, properly gathered, presented, and interpreted, can help provide the appropriate information for goal setting. Perhaps it is here at the very first step of organizing for educational improvement that data-based decision making is most critical and, therefore, must receive careful guidance and consideration. It is at this point that the decision makers must view the entire educational process in terms of their desires. The data under consideration must include and make explicit these desires.

Kinds and Sources of Data. The data necessary for wise decision making must first be identified and gathered. For one thing, demographic data need to be considered so that all those involved in the decision making have a common core of some basic factual information. Included in

this category of information are such things as the number of people in the community, the number of school-aged children, the number of children enrolled in school, the average class size, the rate of truancy, the percentage of children transferred in and out of the school district each year for the past several years, the number of cases of juvenile delinquency, and the number of juveniles involved. Data like these are often already available. Some are in school and/or community offices while other data are found in county and state records. Perhaps all that needs to be done is to gather the information and organize it in a form that would be helpful to the decision makers. For example, identifying the number of children entering school who do not speak English and the number of children in school whose dominant language is not English would be important for making decisions about having, developing, and implementing a bilingual program. In fact, such information would be essential to obtaining special funds for such programs and, therefore, well worth the effort and cost of assembling the data.

Community plans provide another source of information that must be studied. Is the community one that is growing in terms of population? If so, what kind of growth is anticipated? For example, are housing developments and/or apartment complexes planned that will affect the school population and, if so, what is the anticipated effect? If such developments are anticipated, are the developers committed to contributing land for schools and/or community recreation? Are there established community-improvement plans and, if so, how might these relate to the schools? Are there other plans for the community that are relevant to the schools and, therefore, should be considered in making decisions about the educational system?

Information about such community plans is generally available from planning boards and other local organizations and is often helpful for making decisions about school facilities, staffing, and program development and implementation. In addition, data about such things as trends in the labor market, consumer buying, and adult literacy may be available from local and/or state offices and would be helpful. This type of

information about the community, in summarized form, can be used to assist people involved in identifying district and local goals.

For determining the district needs in terms of goals and objectives, the information provided by the demographic data and other data about the community must be augmented by data that are generally obtained only at or through the schools. This would include such things as student medical records, skills-attainment checklists, teachers' observation reports, samples of students' work, records of student- and parent-teacher conferences, students' self-appraisal forms, student and parent questionnaires, students' grades and scores on standardized tests (including diagnostic and statewide measures), students' scores on locally developed measures, supervisors' evaluations of teachers, and records of systematic observations of teachers. Data like these are also useful in evaluating programs.

The data collected at this point in the T & E process are to sufficiently update particulars about the community and the school system so that the decision makers can proceed with goal development that reflects such things as a view of man and society, a theory of learning and development, and a decision about the role of the school. In other words, data collection for a needs assessment as described in Chapter 3 is not what is required for goal setting. Rather, the data should help provide the decision makers with an idea of the community's and educators' philosophy and expectations of education.

Clearly, data gathering for goal setting is not without its problems. Once community pressure groups begin to exert their influence, keeping things in perspective is likely to be a difficult task. At such times, the interests of those other than the pressure groups may easily be swept aside. The most vocal member of the community is often the one who is listened to even when his/her interests or ideas are not generally shared. Just such a situation was illustrated in the example mentioned previously where the goal focused only on the college-bound, ignoring the needs of one-fourth to one-third of the students. Pressure groups should not be ignored, but neither should the interests of others.

Organization of Data for Decision Making. Once the information is gathered, the data need to be organized and presented for a joint and

broad review by the school district's staff and the community so that trends can easily be identified and goals can be set in light of the information presented. Some of these data might be organized and presented to the school district this way:

* = Total	Year			
	1972	1973	1974	1975
Number of students in:	<u>1476*</u>	<u>1466*</u>	<u>1749*</u>	<u>1896*</u>
Grades 1-6	800	811	1020	1100
Grades 7-8	250	245	289	310
Grades 9-12	426	410	440	486
Number of 16-year-olds dropping out in:	<u>2*</u>	<u>7*</u>	<u>13*</u>	<u>16*</u>
Grade 9	0	0	1	0
Grade 10	0	1	1	0
Grade 11	2	4	8	11
Grade 12	0	2	3	5
Number of high school graduates upon graduation:	<u>108*</u>	<u>105*</u>	<u>106*</u>	<u>115*</u>
going to a 4-year college	82	78	53	45
going to a 2-year junior college	10	13	23	29
going to a technical school	3	6	15	20
going to a job	12	6	14	18
unemployed	1	2	1	3

The data above indicate that (a) the school enrollment has increased over the four-year period studied; (b) the increase has been primarily at the elementary school level; (c) there has been a steady increase in the number of high school dropouts, primarily occurring at Grade 11; (d) there has been a steady decrease in the number of high school graduates going immediately to a four-year college; (e) there has been a steady increase of high school graduates going immediately to a two-year junior college or technical school.

These trends from the small sample of data presented above must be examined along with other data, such as information about juvenile crime,

unemployment, and what the community expects of the schools. A problem is likely to develop if a small sample of data, like the one on page 19, is considered in isolation from other relevant information. To avoid such a problem, all data relevant to a specific topic or issue should be grouped together or cross-referenced when presented for consideration by the decision makers.

After all relevant data are carefully studied, judgments must be made about the desired direction of the educational process in the community. Then local goals are set to conform with the state goals.

Of course, since the state goals are applicable to all public schools in New Jersey, it might be easier and perhaps more economical for a school system to begin the local goal-setting process by reviewing the state goals along with local information, including the local philosophy and expectations of education. For example, a state outcome goal is to enable every pupil to acquire the understanding of and the ability to form responsible relations with a wide range of other people, including but not limited to those with social and cultural characteristics different from his/her own. Those setting local goals might review this state goal in light of local data on such things as race relations, number and kind of cultural groups living in the community, languages spoken in the community, socioeconomic levels present in the community, number of aged, housing facilities for retirees, and what the community expects of the schools. Then a local goal or goals could be set to conform to both the locally desired direction of the educational process as well as the state goal.

Evaluation of the Goal-Setting Process

In addition to the setting of goals, there should be an evaluation of the goal-setting process. Evaluation should take place immediately before and after goals are set to be sure that they accurately reflect what is wanted.

An evaluation is needed before goals have been set to determine if the appropriate people and data have been involved or represented in the goal setting. The law identifies the school district's staff, pupils, parents or guardians of pupils, and other district residents as those who

are to be consulted by the board of education which sets the goals. Input from these various groups is obtainable through conferences, meetings, and questionnaires. However, sources for additional data need to be identified by the chief school administrator, who is responsible for developing, for school board approval, the written plan and calendar for the goal-setting process.

Before the setting of goals is begun, the school board needs to ask questions such as these:

- (1) Have available and/or necessary relevant school data been obtained? Does the school staff, including the chief administrator, see omissions that need to be included? If so, how might these be obtained?
- (2) Have available and/or necessary relevant community data been obtained? Who might help answer this question? If additional information is needed, how might it be gathered?
- (3) Have the staff, pupils, parents or guardians of pupils been consulted? Have other district residents been consulted? If not, who has been omitted and how might these people be given the opportunity to provide input?

The second evaluation of the goal-setting process is needed after goals have been set. Here, the questions to be asked might include:

- (1) Were the data sufficient and adequately organized and presented to provide the information needed to set goals? If not, how can (or will) this be remedied?
- (2) Are the goals valid in terms of problems and situations the student will encounter in the world outside of school such as adult performance needed in life situations, the nature of the job market, the changing and expanding nature of knowledge?
- (3) Do the goals conform to the state goals? If not, how can/will this be remedied?

Invariably, there will be problems involved in the goal-setting process. Differences of opinion will be expressed, at times creating

strongly opposing factions. Some members of the community will be pleased with a few of the things done in the goal-setting process; some will be pleased with all things; however, it will be unusual if everyone is pleased with everything.

Summary

In summary, goal setting at the local level must involve educators, pupils, parents, and other community members in order to (a) ensure that the schools continue to be responsive to the needs of all concerned groups and individuals in the community, (b) ensure that local goals are consistent with state goals, and (c) provide a base against which to set objectives and standards for programs and to assess needs.

It should now be clear that goal setting is not an easy task. Points of view as to the purpose and expectations of the school will vary, sometimes causing heated debates. Also, the fact that goals are set may lead people to believe that desired changes in the school will take place immediately. The board of education will have to make clear to the community that a needs assessment based on the goals, objectives, and standards set will be conducted before any changes are made and that the changes needed will be ranked in order of importance and met within the limitations of the budget. Furthermore, the goals first set by the local community may not be lasting. In fact, the legal requirement to set goals every five years is an indicator of the expectation that goals may change just as the world changes.

Setting goals is not going to solve problems. It is only a beginning from which to move forward toward an improved educational process. By setting goals, you attempt to formalize your expectations of the school and provide a basis for determining whether the school is meeting them.

INFORMATION FOR ASSESSMENT AND EVALUATION

CHAPTER 2

OBJECTIVE SETTING AND STANDARD SETTING

Purpose of This Chapter

The purpose of this chapter is to discuss the formulation of objectives, using goal indicators and standards in the context of New Jersey's "thorough and efficient" legislation. The chapter will define the meaning of goal indicators, standards, and objectives; show how objectives are to be related to local goals and needs assessments; discuss ways in which objectives are reviewed and revised; and stress the need for educators' professional judgments throughout the process.

Definition of Objective

As defined by the State Board of Education, "Objective means a written statement of the intended outcome of a specific educational process" (NJAC 6:8-1.1). Obviously, that definition is far from unambiguous, but several elements are clear. The focus is to be on the result or outcome of instruction rather than on the instructional process. Since most education is designed to change students in some way, intended outcomes are likely to be in terms of what students are expected to be able to do (goal indicators) and the degree to which students are expected to do it (standards). If attainment of the objective is ever to be evaluated in some way, the objective should deal with intended outcomes that are observable. Information on how to write and select objectives begins on page 27.

Goal Indicators

The law does not define or require goal indicators, but it is obvious that goal statements must be further specified to be useful. Thus, the first step in formulating an objective is to specify that which students must know if the goal has been attained. For example, spelling might be one indicator of the achievement of a general communications-skills goal.

Definition of Standards

As defined by the State Board of Education, "Standards means the procedures and stated levels of proficiency used in determining the extent to which goals and objectives are being met" (NJAC 6:8-1.1). The definition is extremely inclusive since it covers both the procedures used to measure proficiency and the levels that will be judged acceptable. In effect, the standards are statements of how the objectives will be measured as well as statements of how much is good enough. Standards are discussed in detail beginning on page 32.

Legal Requirements for Objectives and Standards

The first legal requirement to be aware of is that objectives and standards have to be generated. The process is not optional. It must be accomplished by the district board of education in consultation with teaching staff members and the chief school administrator. Objectives and standards must be based on local goals. In addition, under Chapter 97 amending Chapter 212, the state sets minimal standards in some areas of communication and computation skills at selected grade levels.

Clearly, the entire decision-making process will make sense if and only if the steps are logically interrelated. Each objective must be based on, and contribute to, attainment of some goal. If an objective cannot be linked to a particular goal and a rational defense made of the way that objective contributes to the goal, the value of the objective is suspect.

Goal setting, by mandate, is to be carried out by a varied group including educators, pupils, parents, and community residents. Objectives and standards, however, "shall be developed in consultation with teaching staff members under the direction of the chief school administrator" (NJAC 6:8-3.3).

Given the rather technical nature of objectives with standards as opposed to the straightforward nature of goals, the heavy involvement of professionals in the former is necessary. Note that the letter of the law does not preclude--and the spirit of the law encourages--the involvement of other groups such as parents, students, and local organizations.

The procedures involved in setting goal indicators and standards (objectives) require knowledge of curriculum, familiarity with teaching methodology, and experience in the way students are likely to act. The people most likely to have that combination of knowledge, experience, and familiarity are the instructional staff within a school system.

The setting of goals and objectives (indicators and standards) tells you where you want to be. The needs assessment, discussed in Chapter 3, tells you where you are. In combination, they help you determine the way to go.

Setting Objectives

As most of the readers of this document are aware, there is a plethora of texts that explain how to write objectives. Some of the better ones are referenced in Appendix C-2. The texts stress that objectives should deal with observable student behavior. Some texts state that an objective should contain at least four specific elements:

- (1) actor--who is performing
- (2) behavior--what is the performance
- (3) conditions--under what circumstances
- (4) degree--under what standards of correctness

Some texts even provide lists of verbs that are acceptable in writing objectives ("define" is acceptable, "appreciate" is not).

It is not the purpose of this section to replicate those texts but to replace the emphasis on the technical "correctness" of the objective with an emphasis on the meaningfulness and utility of what the student is expected to be able to do.

Types of Objectives. It is possible to subdivide objectives into three major categories: cognitive, affective, and psychomotor. There is a tendency on the part of the lay public to stress cognitive objectives, but most educators are aware of the importance of all three types.

Cognitive objectives emphasize knowing, comprehending, and evaluating. They are, for example, the type of objective required by State Goal 2, "to acquire a stock of basic information concerning the principles of the physical, biological and social sciences . . ."

Affective objectives deal with attitudes, values, and emotions. Though difficult to measure, affective objectives are clearly required by State Goal 12, for example, "to learn to enjoy the process of learning . . ."

Psychomotor objectives focus on physical skills such as the ability to swim or to use a typewriter. State Goal 5 "to acquire job entry-level skills . . ." may require psychomotor objectives.

Though each of the three types of objectives deals with different areas of human endeavor, the same processes may be used to write and evaluate all of them.

Two Questions. For all practical purposes, an objective can be generated by writing the answers to two simple questions:

- (1) What should the student be able to do? (goal indicators)
- (2) How can we tell if the student is able to do it? (standards)

Most, if not all, texts on writing objectives frown on the objective "The student will appreciate music." The texts argue that such a statement is not "behavioral." Perhaps it is not, but many people want students to be able to appreciate music, and it is a perfectly acceptable answer to the first question.

For a response to question 2, get a group of teachers together and ask them how to determine if a student appreciates music. One may say, "Choosing to listen to music in a free-choice situation." Another may suggest knowledge of performers or composers. Yet another may specify the ability to recognize types of music appropriate for different purposes. In a relatively short time, a list of possible ways of telling if a student appreciates music can be generated.

Note that responses to question 2 lead automatically to suggestions for evaluating mastery of the objective. Using the two-question method of generating objectives helps insure the necessary link between objectives (goal indicators and standards). In addition, the ability to answer the first question in natural non-"behavioral" words helps to insure that important objectives are not neglected merely because people cannot think of an "acceptable" way to state them.

A Common Mistake. A mistake often made in the objective-generating process is to produce objectives from teaching materials that happen to be in use such as the lists of objectives publishers often include in their texts. Materials should be selected because they meet objectives rather than objectives selected because they happen to be included in the materials. Publishers' objectives may be a useful aid, but it is the responsibility of the district to select objectives that will be of value to the student.

Utility of Objectives. Except for objectives dealing with basic literacy or other survival skills, it is practically impossible to prove that knowledge of a particular objective will be valuable for a student. Granted the difficulty of empirical demonstration, we still have the responsibility of doing the best we can to write objectives that are worth knowing.

We suggest as a test for the utility of selected objectives a panel of reviewers led by a devil's advocate who should probe and question the rationale given for the inclusion of a particular objective. Such questions should be asked as:

Will it help the student in the next unit of instruction?

Will it help the student outside of school?

Will others think less of the student if the objective is
not mastered?

Will the student think less of himself/herself if the objective
is not mastered?

Which goal is the objective linked to?

Could the goal be met without mastery of the objective?

What would the student miss if the objective were never
mastered?

Can the behavior be observed?

Some objectives, such as those dealing with the ability to comprehend the main idea of a paragraph, will be easy to defend. Others, such as those dealing with the ability to distinguish between simile and metaphor, will be more difficult to defend. And some, such as those dealing with the syntactic distribution of "shall" and "will," may prove impossible to defend. In any case, the bulk of the time that a district spends on writing objectives should be spent on assessing the utility of the objectives and their links to the district's goals rather than in the development of technically "correct" objectives.

District/School and Classroom Objectives. Objectives may deal with results of an instructional process or they may deal with the steps leading up to those results. For example, the ability to read aloud a given passage in a third-grade text may be a school or district objective at the third-grade level. Leading up to that are more specific objectives such as discrimination among shapes, discrimination among letters, or associating sounds with their printed representations.

Since the legislation calls for objectives to be set at the district and school levels, it seems most reasonable to limit objectives at those levels to major bench marks of instruction. The individual classroom teacher may be the one to select the specific objectives that as a result of his/her training and experience are believed to be most appropriate to lead to mastery of the school and district objectives by the particular students in the class. For example, some students may learn to read more efficiently using a sight-word method, while others may learn more efficiently using a phonics method, and yet others may learn more efficiently using some combination of the two. Extremely specific objectives reduce ambiguity, increase precision of measurement, and allow for more detailed needs assessment. But they carry the inherent danger of collapsing under the weight of the paper needed to support them.

Evaluation of Sets of Objectives

Just as each objective in isolation must meet certain criteria of clarity, utility, and level of specificity, the entire set of objectives related to a goal should meet certain criteria.

Sequence. Ideally, the ordering of objectives within a set should be based upon the inherent logic of the content area and supported by well documented theories of learning. For certain content areas, such as mathematics, the ideal may be approached. It is clear that the study of whole numbers should come before the study of rational numbers, that the study of measurement of area should precede the study of the measurement of volume, and that the study of positive numbers should precede the study of negative numbers.

Other content areas, however, are much less likely to have generally agreed-upon sequences based on anything other than tradition. There is still a great deal of argument, for example, about whether or not recognition of whole words should be taught prior to mastery of the set of phoneme-grapheme correspondences.

Given the general lack of empirically verified sequences of objectives, how should the sequence of objectives within a set be established? One reasonable suggestion is to follow hierarchies of the type established by Bloom, et al. in Taxonomy of Educational Objectives, Handbook I: The Cognitive Domain (1956). Theoretically, lower-level objectives need to be mastered prior to mastery of higher-level objectives. A brief example of the suggested sequence within the cognitive domain follows:

- knowledge of specifics
- knowledge of ways of dealing with specifics
- knowledge of universals and abstractions
- translation
- interpretation
- extrapolation
- application
- analysis of elements
- analysis relationships
- analysis of organizational principles
- production of a plan
- derivation of a set of abstract relations
- evaluation in terms of internal evidence
- evaluation in terms of external criteria

Similar lists are available for the affective and psychomotor domains, such as the Taxonomy of Educational Objectives, Handbook II: Affective Domain (1964) by Krathwohl, et al. and The Classification of Educational Objectives: Psychomotor Domain (1966) by Simpson.

Another possible approach to evaluating the sequence of objectives is to study the sequences of objectives generated by recognized experts who publish texts. Radical departures from sequences that are widely used should

be questioned. If rationales exist, however, departures from published sequences of objectives are, of course, acceptable. The point here is that rationales should be given before radical changes in the sequence of objectives are accepted.

Sufficiency. Is the set of objectives complete? Are objectives missing that should have been included? Definite answers to those questions may never be found, but it is possible to check for major omissions.

One approach might be to circulate the list of objectives to the instructional staff and ask if anything is taught that is not covered by the listed objectives. If so, you should consider including these objectives in the list. Of course, care should be given to ensure that additional objectives meet the selection criteria established for previously included objectives. In addition, a review of the textbook material in use should be made to see if they include objectives that are not on the list. Published lists of objectives, such as those available from Instructional Objectives Exchange (P.O. Box 24095, Los Angeles, CA 90024), should also be checked against the district's list of objectives. It is important to remember that objectives should not be added to the district's list merely because they exist elsewhere. Any additional objectives should be reviewed and evaluated using the same standards that were applied to all the other included objectives.

Setting Standards

As defined above, standards clearly are the part of the objectives to be measured and they are the levels of proficiency considered acceptable. Chapter 3 (Needs Assessment) deals with the issue of selecting ways in which the objectives are to be measured and should be used in conjunction with this chapter. This section will be concerned with setting standards. There are three procedural options:

Option I

1. Identify goal indicators.
2. Choose appropriate measures (see Chapter 3).
3. Gather data on each goal indicator using the measures (needs assessment).
4. Examine results.
5. Determine standards (mastery and expectancy levels).

Option II

1. Identify goal indicators.
2. Choose measures (see Chapter 3).
3. Project standards (mastery and expectancy levels).
4. Gather data (needs assessment).

Option III (Combination of I and II)

1. Identify goal indicators.
2. Choose appropriate measures (see Chapter 3).
3. Project tentative standards (mastery and expectancy levels).
4. Gather data (needs assessment).
5. Revise preset standards.

Predetermined standards (Option II) use pure judgment and any data already existing in the district. When you use predetermined standards, you focus on such questions as What will be an acceptable level? or What will be the relationship between one grade level and another? The problem in this approach is that with many of the measures selected there may be no sound basis for judgment.

If you do not use predetermined standards and collect your data before you set your standards (Option I), your judgment will be based on actual data. The questions in this case are these: (1) Are the results adequate? and (2) Are the results appropriate?

The problem with this approach is that you may be tempted to accept what is rather than ask what should be.

Option III is a more complex process but may minimize the problems in both Option I and Option II. Option III might be used in relationship to some of the goals, using Option II where a sound basis already exists and Option I or III for goal areas where little data exist.

Whichever system you use, there are two parts to setting standards at the school and district level: (1) deciding what level of performance is good enough to show proficiency (mastery level) and (2) deciding what percent of the students should be at or above that mastery level (expectancy level). The two parts are obviously interrelated but are different enough to require separate discussion.

Mastery Level

The mastery level is a statement of how well a student has to do on a particular measure or measures in order to be judged a master of the objective that is being evaluated. The mastery level tells how good is "good enough." Setting reasonable mastery levels is quite difficult. Every method yet devised relies heavily on human judgment. This section will discuss several methods of setting mastery levels that are applicable with little or no special training. More elaborate methods requiring statistical expertise will be found in the references cited in Appendix C-2.

The most common method of setting mastery levels involves pooled judgments by professionals using various assessment procedures. In essence, a group of people ask each other, "What seems adequate for a passing performance?" This may lead to the setting of various mastery levels for various groups of students. Though fairly easy to apply, this method may result in inappropriate mastery levels for certain measures. For example, if 80 percent is selected as a mastery level, it may be fine for one measure; but if another measure of the same objective happens to be much more difficult, many students

demonstrating mastery on the first measure may fail to show mastery on the second.

Another method requires a group of judges who are familiar with students. The steps to be followed are:

- (1) Ask each judge to think of the lowest level of performance that (s)he is still willing to classify as minimal mastery level on the given objective. Many judges are helped in their conceptualization of the lowest level of performance still acceptable as mastery by thinking about the performance of particular students. The intent is not to institute a norm-referenced performance standard but to help judges establish a criterion-referenced standard.
- (2) Look at the first task on the measure and try to decide the probability that the worst performing master would get the question right or be able to perform the task. If the task looks easy, there is a high probability the worst performing master would answer the question correctly. If the task looks hard, there is a low probability. If the judges have difficulty with the estimation of probability in the abstract, ask them if they can conceptualize 100 students performing at the minimal mastery level and to state the number out of 100 that they would expect to get the task right. Such a process is equivalent to estimating probabilities directly.
- (3) For each task, calculate the average probability estimated by the judges.
- (4) Add up all the average probabilities. (There should be as many averages as tasks in the measure.)
- (5) Use the sum of the probabilities as an estimate of the mastery level.

The following example has been worked out for five judges and a ten-task measure:

SETTING A MASTERY LEVEL BY JUDGES' ESTIMATES

Task	Judges' Estimates					Average Probabilities
	A	B	C	D	E	
1	.87	.68	.93	.77	.90	.83
2	.76	.60	.85	.69	.85	.75
3	.77	.65	.87	.72	.87	.78
4	.59	.43	.50	.65	.70	.57
5	.62	.51	.65	.68	.70	.63
6	.75	.80	.77	.71	.78	.76
7	.25	.30	.15	.20	.35	.25
8	.0	.05	.10	.10	.07	.06
9	.33	.30	.35	.40	.37	.35
10	.29	.45	.47	.40	.33	.38
						5.36 = sum = mastery level

According to the example, a score of 0 to 5 would be below the mastery level. A score of 6 to 10 would be above the mastery level.

For those areas of the curriculum that are sequential (mastery of step 1 is prerequisite to mastery of step 2), the following method of estimating mastery levels is appropriate:

- (1) Ask each judge to look at each task and try to estimate the probability that the skill measured by the task will be required in the following unit of instruction. Clearly, the judges have to be extremely familiar with the curriculum in order to be able to make the required judgments.

- (2) Follow steps 3-5 on page 35.

Elimination of Choices. This method is limited to multiple-choice questions. Again judges are used who are familiar with both students and curriculum. They are asked to think of the lowest level of acceptable performance as in the method described previously and are asked to look at each question in turn. The steps to be followed are:

- (1) Decide which choices in the multiple-choice question a person performing at the lowest level of mastery would be able to eliminate as clearly wrong.
- (2) Count the remaining choices in the question.
- (3) Divide 1 by the number of remaining choices.
- (4) Treat the result as a probability and continue with steps 3-5 as described for the method on page 35.

Consider the following sample question:

Newark is a

- (A) city
- (B) state
- (C) county
- (D) continent
- (E) world

If a judge indicates that even the worst performing master would be able to eliminate choices C, D, and E as wrong, that would leave two choices. Dividing 1 by 2 gives .5 which is the probability that a worst performing master will be able to answer the item correctly.

Empirical Methods. The methods of setting mastery levels discussed above all depend directly on the judgments of a group of people. Several empirical methods based on data derived from student performance on the measure are available. Some techniques require knowledge of regression analysis or Bayesian statistics and are beyond the scope of this chapter. A method that requires no more statistical sophistication than the ability to compute averages is described below. Note that all methods of setting mastery levels require human judgment at some stage in their application.

- (1) Select (by some means independent of the test or measure on which a mastery level is being set) a group of students judged to be nonmasters of the objective under consideration.
- (2) Select (by some means independent of the test or measure on which a mastery level is being set) a group of students judged to be masters of the objective under consideration.
- (3) Ascertain that the groups are as similar as possible except for mastery of the objective under consideration.

- (4) Administer the test to both groups.
- (5) Calculate the mean score of each group.
- (6) Calculate a point midway between the two mean scores.
- (7) The midway score is a reasonable estimate of the mastery level.

Expectancy Level

An immediate reaction one usually has to the question of how many students should be expected to master each objective is 100 percent. But is this realistic? When you consider the differences among individual students in terms of their rate of growth--physical, emotional, and intellectual--as well as the order in which new abilities are acquired--it is not realistic to expect 100 percent of the students to achieve each objective. Nor is it realistic to expect all students to achieve each objective at the same level of mastery and at the same time.

An expectancy level should be treated as a tentative hypothesis and reviewed at least once a year as new information becomes available. But you must start somewhere to develop your first stated hypothesis. Perhaps the best way to begin is to ask yourself: For this objective, what level of mastery can we reasonably expect our students to attain? And to help answer this question, ask another: For this objective, what level of mastery is expected to be attained by the entire population? The answer to this latter question helps put things in the proper perspective. Some data are available from such things as the National Assessment Study, national norms on various testing programs, and requirements for various kinds of jobs and other activities of everyday life.

It is not necessary to set mastery expectancy levels at each grade. A logical cut-off point can be identified. For example, cut-off points of mastery of the objective of adding whole numbers might logically be set at third, sixth, and ninth grades. In this way, progress over time can be monitored and used to revise expectancy levels if necessary. Then, by expressing the expectancy levels in terms of the percent of the student population being considered, you allow for the individual differences discussed above.

In setting expectancy levels, it is important to remember that you must consider the characteristics of the student population for which the levels are being set. Are the students in high school, middle school, or elementary school? Are they non-English-speaking students in an English-speaking class/school? Are they exceptionally gifted and talented students and, if so, in what way(s)? Are they handicapped students? Try to set expectancy levels that are realistic for the student population being considered and revise them if necessary.

Grade-Level Expectancies. The expected level for most objectives should increase with each grade level. If, for example, an objective dealing with the ability to distinguish between fact and opinion is introduced in grade 3, many students may not be expected to master it. But the objective is reinforced in subsequent years. It is reasonable to set increasing expectancy levels with increasing grades for those objectives that are taught across a span of years.

Note, however, that for those objectives not reinforced over time, learning theory and teaching experience both indicate that student competence will decrease over time. To be realistic, expectancy levels for nonreinforced objectives should be set lower with increasing grades. If setting reduced expectancy levels is seen as objectionable, then reinforcement of the objectives should be added to the curriculum.

Expectancies Change. An expectancy level should be treated as a tentative hypothesis. Expectancies ought to be reviewed at least once a year as new information becomes available. For example, a district may set an 85 percent expectancy level for an eighth-grade objective dealing with knowledge of the effects of smoking on health. The needs assessment may show that 97 percent of the students master the objective. If consultation with administrators, curriculum supervisors, and instructional staff show that none feels an inordinate amount of time is being spent on the objective, the expectancy level may be increased.

If, on the other hand, an expectancy level has been set at 80 percent and the needs assessment indicates 57 percent mastery, the resources allocated to the objective may be increased significantly. The next year's

needs assessment may show 71 percent mastery, but all staff involved may feel that the objective is worth no more time than is being given. Consideration should be given to lowering the expectancy level unless people are willing to allocate even more resources to the objective.

INFORMATION FOR ASSESSMENT AND EVALUATION

CHAPTER 3

NEEDS ASSESSMENT

Purpose of This Chapter

The major purpose of this chapter is to explain how to conduct a needs assessment in the context of New Jersey's "thorough and efficient" legislation. The chapter will define the meaning of "needs assessment;" indicate the requirements specified by law; show how to plan for a needs assessment; identify selected measures--both test and nontest measures--as related to the state goals; show how the data collected can be used to identify needs; and stress the need for educators' professional judgment throughout the process.

Definition of Needs Assessment

As defined by the legislature, "needs assessment means a written analysis of the current status of an educational system in terms of achieving its goals" (Article I, Chapter 212). Essentially, such an assessment is carried out to determine where you are as related to where you want to be. A needs assessment helps you find out what needs to be done to meet your goals. The difference between the level of achievement desired and the present level of achievement, determined from the needs assessment, will form the basis for new or modified educational programs.

Legal Requirements for Needs Assessment

In accordance with the Public Education Act of 1975, each school district is to make an annual report of its progress in conforming to the goals, objectives, and standards developed for a "thorough and efficient" process of education. Such reports are to include, among other things, results of each school's effectiveness in achieving both state and local goals and the related objectives applicable to the pupils (Article II, Chapter 212).

Furthermore, the administrative code concerned with the "thorough and efficient process of education" legislation requires that needs of students be assessed by teachers to determine if the students have attained both the long- and short-range objectives specified by the school district. Moreover, more than one assessment technique on which to evaluate the attainment of objectives should be utilized to provide as complete and adequate an assessment as possible. Assessment procedures are to include, but not be limited to, such things as teacher-observation reports, parental or guardian interviews, cumulative pupil records, medical examinations, formal and informal education techniques, and local and state testing results (NJAC 6:8-3.4). It is important to note that test results alone are not sufficient.

Planning for Needs Assessment

In order to obtain the evidence needed to make judgments about whether or not goals and the related objectives are being attained and to what degree they are being achieved, you must plan carefully. The information required for such an analysis must be both sufficient and valid.

Need for a Broad Data Base. It is essential to use more than one way to obtain the data. Assessments like this are too often limited to standardized, published paper-and-pencil tests because they tend to be objective and dependable, economical in terms of time, and easy to administer. Moreover, they often provide national norms against which to make comparisons, they may provide scores that can be compared with local standards of achievement, and because each test is available in several forms, a systematic measurement of progress may be possible. However, tests have their limitations. Objective tests, which are designed primarily for measuring factual knowledge, are not capable of measuring many kinds of objectives. Even with the variety of tests available, it is often difficult to find tests that adequately measure objectives in the way in which a school district wants them measured. If the ability to write essays is an objective, for example, the most valid way to measure this is to have students write essays. Furthermore, several samples of essay writing from each student would be helpful since a

student's performance one day may be quite different from what it is on another day due to a variety of things ranging from health to interest in and knowledge about the topic of the essay question. Although writing ability is frequently measured by multiple-choice objective tests, an adequate and valid assessment of this skill should include samples of student's writing.

Thus, evidence from several sources is needed to make judgments about whether or not objectives are being achieved and to what degree. Furthermore, for each objective, you need information from more than one measure since any given measure is only a sample of what students can do in relation to that objective. And, like the essay writing indicated above, the information obtained may be affected by a variety of things. Therefore, the more pieces of information you have about students' attainment of an objective, the more certain you can be about deciding whether or not it is being accomplished and, if not, what needs to be done.

Suppose, for example, that you wanted to assess the students' attitudes toward human relations. To do this, you decide to use The Human Relations Test (Test 3 of the Educational Goal Attainment Tests) and two essays requiring students to resolve given problems in human relations. Many students know the kind of responses that are looked upon as appropriate and may, therefore, respond accordingly although their attitudes are not really what the test and essay results indicate. For this reason, more information would be helpful in determining the students' real attitudes. Such data may be obtained from teachers' observation records, discipline records, and other sociometric devices. The more information that is available, the more accurate and valid will be the assessment. However, you should keep the time and resources needed for data collection within reason.

Identification of What Is To Be Done. You should carefully plan your needs assessment in terms of the time and resources that appear to be available both currently and over time. Because a broad data base is needed to do a good assessment and because time for assessment is limited, a schedule for implementing the plan should be devised. Remember, if you try to measure everything at one time and in great detail, there will be no time for instruction. Ask yourself at each step of the plan: "What will this tell me about what I need

to know?" and "Will the knowledge gained be worth the effort?" If not, modify your plan.

To obtain sufficient, adequate, and valid information for decision making, the needs-assessment plan must include the school district's goals and objectives along with its corresponding standards. In addition, the samples on which data are to be collected need to be specified. (There is a discussion about identifying a sample in Appendix B-1.)

On the following pages, you will find three examples from needs-assessment plans, each of which has explanatory notes at the end of this chapter. For each objective related to a given goal, the following is identified:

- (1) The sample on which the data are to be collected
- (2) The measure to be used in obtaining the data
- (3) The expected level of proficiency for each measure

Identification of the Sample. Since the purpose of an assessment is to see how all the students are achieving the school district's objectives, it is not necessary to collect information on every individual. However, if you do not plan to include everyone in the data gathered, you must define the sample carefully to prevent bias. That is, those on whom data are collected for making decisions about a larger group should be representative of, and have the same characteristics as, that larger group. For example, if you wanted to assess the achievement of science skills by eighth-grade students, you would not have to include data on all eighth-grade students in a school district. Ideally, you would select a stratified random sample of students across the district. This would require identifying various strata in the population (e.g., based on sex, race, age, home or school environment, or combinations of any of these) and then randomly selecting a sample of individuals from each stratum that would reflect its proportion of the total population. If this created scheduling problems, perhaps you could obtain adequate data from one-fourth of the eighth-grade classes in each of the schools in the district. Of course, the classes chosen should contain a sample of students that is representative of all eighth graders in the school and/or district, including such characteristics as age, sex, race, and possibly home and school environment. Given these characteristics and the variability

EXAMPLE A FROM A NEEDS ASSESSMENT PLAN¹

GOAL	GOAL INDICATORS	STANDARDS		SAMPLE
		EXPECTANCY LEVELS	MEASURES	
To acquire knowledge, habits, and attitudes that promote personal and public health, both physical and mental	Students will <u>not</u> smoke cigarettes.	100% below 8th grade 95% grades 8-9 90% grades 10-11 85% grade 12	Student questionnaire	all
			Number of students disciplined for smoking on school grounds	all
			Conference with parents	5% random sample in each grade
	Students will demonstrate knowledge of basic nutrition.	75% grade 8 85% grades 9-12	Given a menu from a local restaurant, the student will select one breakfast, one lunch, and one dinner that in combination contain at least one food from each of the basic groups.	25% random sample from each grade 
			Given a list of 20 common foods, the student will select 3 of the 5 that are highest in protein.	
			The student will name at least 3 sources of each of the vitamins A, B, C, and D.	
	Students will be able to apply basic first aid procedures.	60% grades 8-9 75% grades 10-11 90% grade 12	The student will be able to locate at least one pressure point to stop arterial bleeding for each extremity.	25% random sample from each grade
			The student will list the steps to be followed in treating a victim in shock including at least 4 appropriate procedures and <u>NO</u> counter-productive procedures.	

- 47 -

EXAMPLE B FROM A NEEDS ASSESSMENT PLAN²

GOALS	GOAL INDICATORS	STANDARDS		SAMPLE
		LEVELS OF PROFICIENCY	MEASURES	
To develop an understanding of one's own worth, abilities, and limitations	By seventh grade, students will demonstrate a feeling of self-worth.	85% of all seventh-graders will score a minimum of 60 on the <u>Educational Goal Attainment Tests: Self Test Cluster</u> using subtests 1A, 1B, and 1C.	<u>Educational Goal Attainment Tests: Self Test Cluster</u> , 1A, 1B, 1C	All seventh-grade students in the following classrooms: <u>School</u> <u>Homeroom</u> Middle School No. 1 7A, 7B, 7C Middle School No. 2 7A, 7B The sample consists of: one-third of all seventh graders; a representative sample in terms of age, sex, race, grades (based on student records), and neighborhood characteristics within the school district.
		85% of the seventh-graders will present a positive picture of self-worth, although they may also acknowledge shortcomings.	Student-teacher conferences	Same as sample above
		80% of the seventh-graders will write autobiographical essays that present a positive picture of self-worth.	Autobiographical essays in which students are asked to list achievements, skills, interests, etc.	Same as sample above

EXAMPLE C FROM A NEEDS ASSESSMENT PLAN²

GOALS	GOAL INDICATORS	STANDARDS		SAMPLE										
		LEVELS OF PROFICIENCY	MEASURES											
To acquire basic skills in reading, writing, and mathematics	By fourth grade, students will be able to identify the main idea and supporting details as stated in material read.	85% of all fourth-graders will be able to correctly answer a ratio of 4 out of 5 questions related to this objective on the <u>New Jersey Educational Assessment Program Reading Test</u> .	<u>New Jersey Educational Assessment Program Reading Test</u>	All fourth-grade students —whole population tested										
		80% of the fourth-graders will be able to correctly respond to all questions on an informal reading inventory related to this objective.	Informal reading inventory in which students will be asked, after silent reading of a page or less, to read orally the parts that answer specific questions (i.e., what is the main idea, read something that helps support or prove the main idea, etc.).	All fourth-grade students in the following 10 classrooms: <table><tr><th><u>School</u></th><th><u>Room Numbers</u></th></tr><tr><td>Martin Luther King</td><td>101, 104</td></tr><tr><td>Thomas Jefferson</td><td>211, 212, 214</td></tr><tr><td>Jane Adams</td><td>106, 109, 111</td></tr><tr><td>John F. Kennedy</td><td>110, 114</td></tr></table> The sample consists of: one-half of all fourth-graders; a representative sample in terms of age, sex, race, grades (based on student records), and neighborhood characteristics within the school district.	<u>School</u>	<u>Room Numbers</u>	Martin Luther King	101, 104	Thomas Jefferson	211, 212, 214	Jane Adams	106, 109, 111	John F. Kennedy	110, 114
		<u>School</u>	<u>Room Numbers</u>											
Martin Luther King	101, 104													
Thomas Jefferson	211, 212, 214													
Jane Adams	106, 109, 111													
John F. Kennedy	110, 114													
80% of the fourth-graders will have <u>no</u> mention of problems on this objective in the teacher's anecdotal records.	Teacher's anecdotal records based on classroom observations	Same as sample in 2 above.												

that is possible among groups of students, it is easy to see why more than one measure is needed to determine achievement of an objective.

(For some rules of thumb on selecting a sample and interpreting sample data, see Appendix B-1. Also, information is available in some of the texts referenced in Appendix C-2.)

Selecting Measures

Remember that the results of any assessment consist of data on which to base further study. Moreover, results must be coordinated with background facts and they must be verified by comparison with other available data. It is for this reason that multiple measures should be used in any assessment. Furthermore, prior to selecting any measure, whether a paper-and-pencil test or performance measure, you must know the following about the assessment situation:

- (1) The purposes of assessment (i.e., goals, objectives, and standards)
- (2) The characteristics of the students involved (e.g., age and/or grade level; language spoken, read, and written; etc.)
- (3) The kinds of decisions to be made based upon the interpretation of the results (i.e., continuation of current practice or program modification or new program selection and implementation)

Once you know the above, you can review measures and determine their appropriateness to the situation. It is critical to know each measure completely: how it measures what it is said to measure and all available information on its past use including any statistical data available. Further, since the measures will provide indicators of goal attainment, you will need to carefully examine the indicators you expect to obtain. Following is a checklist of questions to ask yourself about each indicator:

- (1) How relevant is the indicator to the goals, objectives, and standards set?
- (2) Is the indicator reliable?
- (3) How many factors affect the indicator? Do you know which effects are produced by each factor? Do you know why?

- (4) Who is to gather information for the indicator? Are these people properly trained and appropriately placed?
- (5) Is the indicator subjectively or objectively quantified?
How does this affect it?
- (6) What effects, if any, does the indicator have on the populations or institutions it assesses/evaluates? (For example, are there negative emotions or reactive effects that can confound interpretation of the data?)
- (7) How vulnerable is the indicator to response patterns (i.e., responses that are not thought through)?
- (8) Is the indicator biased against or in favor of specific populations? If so, how?

Essentially, two kinds of measures can be considered: internal measures, which are developed within a school district, and external measures, which are available from sources outside the school district.

Internal Measures. There are many types of internal measures. These range from teacher-made observations and tests in a given curriculum area or a specific unit of study to attitude measures, observation scales, sociograms, and other measures that are carefully designed and developed by committees of teachers, supervisors, guidance counselors, and other professional staff members and appropriate interested parties.

The process of constructing an internal measure is complex and time consuming. Nevertheless, such measures are often worth it because they generally are specifically related to local objectives and, therefore, may be more valid than an external measure. In fact, for some purposes only internal measures such as those providing indicators of observed study habits, absence from school, achievement on daily assignments, or class participation may be appropriate.

Test developers themselves must meet certain requirements, the first of which is knowledge of students, human growth and development, and knowledge of what is to be assessed. They must also be able to describe specific objectives for the instrument and be able to effectively communicate what is wanted or being measured. Finally, the developers must be able to use various formats to collect different kinds of information such as whether students can

function at various levels of knowledge like recognition, recall, translation, interpretation, extrapolation, application, analysis, synthesis, and evaluation.

In deciding whether an internal measure should be developed, it is advisable to:

- Search the literature to determine if there is a published measure available that will tell you what you want to know. If there is none,
- Decide if the information to be obtained from developing one will be worth the time and money it will cost.

Assuming that a measure must be developed, the steps within this procedure generally are to:

- (1) Establish specifications
- (2) Produce materials to be included
- (3) Try out the materials by having others review them
- (4) Assemble the materials into the final draft
- (5) Try out the assembled measure and review it if necessary
- (6) Determine how good the measure is

Before constructing test materials, the developers must determine the nature of the population being tested and the purposes for testing. Once these things are known, test specifications can then be set. The test specifications should include: the specific objectives, content, and/or skills to be measured and the weights (in terms of importance) to be given to the objectives, content, and/or skills. If all areas of the specifications are to be equally weighted, this should be indicated.

Next, the kind of test question(s) to be used must be identified. The type of question selected should be one that gets the information wanted, as indicated by the specifications. Each item, regardless of the type, must clearly state the task or problem that the students must react to, complete, or solve. (A discussion of various item types and their uses may be found in some of the references in Appendix C-2.)

After the items have been assembled into a measure, they should be tried out. The appropriate revisions can then be made. Once the internal measure

has been developed, you should determine how good it is so that you will know that any data obtained by it will be valid and reliable.

One reason published measures are usually considered to be of good quality is that they typically go through statistical analyses for item selection and overall evaluation. (Information about statistical analyses is available in some of the references in Appendix C-2.) However, there are simple techniques that you can use to obtain the same kind of information. Paul B. Diederich's monograph, Short-Cut Statistics for Teacher-Made Tests, published by Educational Testing Service in 1973, is an excellent reference for this. Appendix B-2 contains some information from this publication which you may find useful.

Two important factors on which you should judge your internal measure (or any other way of estimating needs) are its validity and usability. To determine its validity (whether it measures what it is supposed to measure), you might have it reviewed by judges who can provide an opinion about its accuracy and adequacy for measuring the objective. You might also administer another measure of the same objective and compute the correlation between the two, using the method described in the discussion on reliability in Appendix B-2. The higher the correlation, the greater the similarity in what they measure. In addition, you can--and should wherever possible--compare results on the measure with the actual performance of the student on the objective or the thing being measured--for example, his or her daily work in mathematics--through observation, daily assignments, checklists, or other means. In other words, a measure is validated by information from other measures of the same thing. Therefore, with good reason, more than one measure is required to evaluate the attainment of an objective (NJAC 6:8-3.4).

To determine a measure's usability, you need only ask yourself: Does it provide data that are both meaningful and usable for my purposes? Whatever your internal measure is--a standardized discipline record, an essay or multiple-choice test at the end of a course, a multiple-rater evaluation of art work, grades in a course, a standardized checklist of school staff by areas of responsibility--you must be concerned about its validity and usability.

External Measures. Just as there are many types of internal measures, there are also various kinds of external measures. The best-known of these are standardized achievement and aptitude tests. However, there are many other kinds of external measures, including such things as questionnaires, essays, observation scales, and checklists. Standardized instruments generally measure achievement, aptitude, and personality. The main difference between standardized measures and others is that the former have been administered uniformly to a representative group of students for the purpose of establishing norms. Generally, a standardized measure covers a broader range of content than an internal one, and its development usually involves the combined efforts of teachers, curriculum specialists/psychologists, and measurement experts.

Criterion-referenced measures are deliberately constructed to provide information that is directly related to given objectives with specified performance standards. Some standardized, norm-referenced measures may also be criterion-referenced or used as such if there are specified performance standards or cut-off scores related to specified objectives.

In general, uses for norm-referenced and criterion-referenced measures are as follows:

Norm-referenced measures

- To rank-order individuals
- To compare an individual with others or himself/herself
- To measure broad areas of knowledge

Criterion-referenced measures

- To assess attainment of objectives
- To describe an individual's performance as related to a goal
- To diagnose problems

Both

- To predict later performance
- To evaluate progress

Criterion-referenced/objectives-referenced measures are generally based on objectives that are much more limited in scope than those of a norm-referenced measure. Further, a criterion-referenced instrument generally has

many more items for each objective than does a norm-referenced instrument. For example, a norm-referenced measure with sound-symbol correspondence listed as one of its specifications might have two items to measure this broad topic. The specifications of an objectives-referenced measure, on the other hand, might list as an objective the ability to associate initial consonant sounds with their graphemic equivalents and include three to four items to measure this. Then, if a criterion is to be established that represents a passing level, it is generally set in terms of students' being able to pass a certain number of items out of the total number of items on a given objective.

Caution, however, is still needed even when there are many items in an instrument to measure a given thing. In general, it is almost impossible to be able to measure the universe of an objective given all the situations in which that objective might be performed. Clearly, even 10 items would not be sufficient to measure all cases of associating initial consonant sounds with the graphemic equivalents. Diagnostic measures are also limited in the same respect although they may be much more complete than other types of measures.

Robert Ebel points out in his article "Educational Tests: Valid? Biased? Useful?" in the Phi Delta Kappan (October 1975) that criterion-referenced/objectives-referenced measures treat learning as if it were acquired by adding separate, independent pieces to the collection of what is learned, whereas norm-referenced measures follow the notion that learning consists of building a structure of interrelated parts. He adds that criterion-referenced measures may create special problems:

- (1) Selecting and defending a unique set of abilities that each student is expected to learn
- (2) Rationally defining a particular level of performance that indicates attainment of each objective
- (3) Repeated testing of those who do not reach the criterion level at first, with the problem of having multiple forms available
- (4) Reporting only two levels (pass or fail) of achievement on

an objective such as reading proficiency that exists at many different levels

- (5) Producing, using, and filing detailed bulky reports for each student

Norm-referenced measures, on the other hand, may also create problems, particularly for those who need special educational programs (e.g., compensatory programs). For these students, pre- and posttest results from norm-referenced tests may not show growth that may indeed have occurred. It may take as long as three to four years before growth is indicated with a norm-referenced measure simply because the instrument is not refined enough. In cases like this, you may need to refine it further or use criterion-referenced/objectives-referenced measures that match the goals and objectives of the program. Of course, any such criterion-referenced measures would need to focus specifically on what is being taught, and if the objectives were too broadly stated and measured, they too might not immediately show growth.

Again, for selecting a measure, it is critical to know why you are measuring, whom you are measuring, and what decisions will be made based on the results.

The following discussion of various kinds of measures is based primarily on Lecture Notes for a Testing Workshop, an article written by S. Valentine and M. Zieky and published by Educational Testing Service in 1976. For purposes of discussion, we are presenting classifications of measures in contrasting pairs:

- (1) Objective--Essay
- (2) Aptitude--Achievement
- (3) Psychometric--Impressionistic
- (4) Group--Individual
- (5) Paper-and-Pencil--Performance
- (6) Cognitive--Affective

Each of these is discussed briefly below. Keep in mind that these classifications are arbitrarily set and do overlap (i.e., an objective measure may also be a paper-and-pencil measure).

(1) Objective--Essay Measures

The basic distinction is that in multiple-choice or objective testing, the most important person is the item writer; (s)he must do a good job or the test is worthless. As its name indicates, anybody can score an objective test and do it correctly. In the essay test, the most important person is the test scorer and if (s)he does not do a good job, the test is worthless. In multiple-choice measures, the student selects a response that is already prepared whereas in essay tests the student provides the response. The former may be considered a recognition operation, while the latter is strictly recall. However, in both cases, the student may be required to interpret, extrapolate, apply information, analyze, synthesize, and/or evaluate.

(2) Aptitude--Achievement Measures

The dichotomy between these two kinds of measures is not strict, but the different purposes are clear. The aptitude test looks ahead and attempts to estimate how well you will do if you are exposed to some kind of instruction. The achievement test looks backward and attempts to estimate how well you have done after you have been exposed to some kind of instruction.

The problem is that there is no way to measure aptitude directly; you must use some kind of achievement. An example of an aptitude measure is the Scholastic Aptitude Test (SAT), which is used to estimate how well students will perform in college. To do so, it measures achievement in mathematics, vocabulary, and reading comprehension. When using aptitude tests, one assumes that all who take them have had an equal opportunity to learn the necessary material, but this assumption is not true. We know, for example, that the SAT is not a fair test of aptitude for Puerto Rican students who use English as a second language because they have not had an equal opportunity to learn English with a native speaker of English. For

a Spanish-speaking person, the SAT is a test of achievement in English, not aptitude for college work.

(3) Psychometric--Impressionistic Measures

Again, this is not a strict dichotomy. The difference is related to the objective-essay difference discussed above. The psychometric test is objective and yields a score which is reported as a number. A correct or appropriate response must be indicated for such tests in order to report scores.

In an impressionistic test, a score is sometimes reported, but the tester concentrates and reports on the way the person taking the test reacts, how (s)he does what (s)he does (e.g., is anxious, or self-assured; impulsive or hesitant; cooperative or uncooperative; talkative or quiet). The problem of impressionistic testing is like the problem of essay testing: You have to put a lot of faith in the person giving the test; his/her judgment might change over time, and different judges do not always agree with each other. Yet the impressionistic testers say that the psychometric testers throw away the most important information.

(4) Group--Individual Measures

The split here is fairly obvious. The group test is cheaper because it is given to a lot of people at one time, but you cannot tell if all members of the group are interested in the test and are working on it. Also, you can't tell if all members of a group understand the directions and are answering in the proper way.

When testing one person at a time, you can be certain about what (s)he is doing at every moment. Individual tests are necessary with young children and with special groups of adults such as the retarded. Individual testing is probably best, but there is not enough money and there are not enough people who know how to do it to use individual tests very often.

(5) Paper-and-Pencil--Performance Measures

In a paper-and-pencil test, the student answers questions about what (s)he would do in a particular situation. In a performance test, the student does it. Performance tests are very important in many fields, but we do not use them as often as we should or would like to because they are not as convenient as paper-and-pencil tests. Measures that simulate actual performance have been developed. For example, in a paper-and-pencil measure of ability to perform a specific chemical experiment, each student may proceed through a different set and/or order of questions depending upon his/her responses. One result of such a test is a successful experiment while the others are unsuccessful, one possibly even culminating in an explosion (still on paper, of course).

(6) Cognitive--Affective Measures

Cognitive tests measure what people know while affective measures help determine how people feel: their attitudes, feelings, and beliefs. The problem with affective tests is that often people want to please the examiner and do not admit feelings, beliefs, or attitudes that they do not think are proper. If a person hates his/her school, (s)he will very rarely admit it, because we are supposed to like our schools. So if we just ask a multiple-choice question like this:

How do you feel about your school?

- (A) I like it very much.
- (B) It is OK.
- (C) I do not care about it.
- (D) I dislike it very much.

how many people would answer C or D, if these were their true feelings? Is there a way, then, to measure attitudes, beliefs, and feelings? Generally, it is best to ask about behaviors that relate to the topic in mind such as: "How often are you absent from school?" or "When is the last time you went to a

school function?" or "Would you like to see your school change in any major way? How?" Generalizations or inferences can then be made based on the answers to these questions.

Descriptions of some specific external (published) measures are found in Appendix A-3.

Identifying Measures Related to State Goals

Some types of nontest measures (both external and internal) that may be used for the various state outcome goals are indicated in the matrix on page 61. There is more detailed information about measures and goals in Appendices A-1 - A-3. Appendix A-1 illustrates in a matrix how 243 instruments are matched with New Jersey State outcome goals. Appendix A-2 provides some basic information about characteristics of the instruments in Appendix A-1. Appendix A-3 has greater detailed descriptions of many of the measures listed in Appendix A-1 and A-2.

While the information here should help you in the selection and/or development of measures related to state goals, you may encounter difficulties. State outcome goals 1, 2, 4, 5, 8, 9, and 12 can be measured to a certain extent by paper-and-pencil measures; however, even these as well as other goals require other kinds of measures, such as observation techniques, questionnaires, and direct-performance measures. Further, the attainment of a goal like 3, "to become an effective and responsible contributor to the decision-making processes of the political and other institutions of the community, state, country, and world," cannot be fully determined for most students since, before they leave school, there are only a few opportunities available to demonstrate goal attainment. Attitude measures, observation scales/checklists, and measures that use simulated experiences are possibilities to be used for predicting attainment of such a goal. Also, attainment of a goal like 11, "to develop an understanding of his/her own worth, abilities, potentialities and limitations," can seldom be directly measured with great assurance as to accuracy and validity. For this kind of a goal, an attitude measure along with some observational measures might be best.

New Jersey State Outcome Goals

✓ = Likely source of evidence

1. To acquire basic skills in obtaining information, solving problems, thinking critically and communicating effectively.
2. To acquire a stock of basic information concerning the principles of the physical, biological and social sciences, the historical record of human achievements and failures and current social issues.
3. To become an effective and responsible contributor to the decision-making processes of the political and other institutions of the community, state, country and world.
4. To acquire the knowledge, skills and understanding that permit him/her to play a satisfying and responsible role as both producer and consumer.
5. To acquire job entry level skills and, also, to acquire knowledge necessary for further education.
6. To acquire the understanding of and the ability to form responsible relations with a wide range of other people, including but not limited to those with social and cultural characteristics different from his/her own.
7. To acquire the capacities for playing satisfying and responsible roles in family life.
8. To acquire the knowledge, habits and attitudes that promote personal and public health, both physical and mental.
9. To acquire the ability and the desire to express himself/herself creatively in one or more of the arts, and to appreciate the aesthetic expressions of other people.
10. To acquire an understanding of ethical principles and values and the ability to apply them to his/her own life.
11. To develop an understanding of his/her own worth, abilities, potentialities and limitations.
12. To learn to enjoy the process of learning and to acquire the skills necessary for a lifetime of continuous learning and adaptation to change.

Nontest Measures													
Questionnaires and interviews	Logs and diaries	Classroom observations	Teachers' observations	Performance rating scales	Teachers' anecdotal records	Self-appraisals	Peer evaluations	Student-teacher conferences	Parent-teacher conferences	Samples of daily work	Skills checklists	Cumulative records	Social indicators
✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	✓	✓	✓	✓	✓	✓		✓		✓	✓	✓	
	✓	✓	✓	✓	✓	✓	✓	✓	✓		✓		✓
✓	✓	✓	✓	✓	✓		✓	✓	✓		✓		
✓	✓	✓	✓	✓	✓	✓		✓	✓	✓	✓	✓	
✓	✓	✓	✓	✓	✓	✓	✓	✓	✓				✓
✓	✓	✓	✓	✓	✓	✓	✓	✓	✓				✓
✓	✓	✓	✓	✓	✓	✓	✓	✓	✓				✓
✓	✓	✓	✓	✓	✓	✓	✓	✓	✓				✓
✓	✓	✓	✓	✓	✓	✓	✓	✓	✓				✓
✓	✓	✓	✓	✓	✓	✓	✓	✓	✓				✓
✓	✓	✓	✓	✓	✓	✓	✓	✓	✓				✓
✓	✓	✓	✓	✓	✓	✓	✓	✓	✓				✓

Nevertheless, another question arises: Should an individual's understanding of his/her own worth, abilities, potentialities, and limitations be evaluated? Do you need to know if the individual's perceptions are inflated or overly modest? If you are concerned with accurate measurement, yes. If you are concerned with valid measurement (that the measure tells us how one perceives oneself), no. For accurate measurement, you may be able to use independent judges to verify individuals' self-perception. Thus, by careful planning and instrument selection and/or development, you can anticipate the difficulties and determine possibilities for assessing attainment of state outcome goals.

You may also encounter difficulties in assessing the attainment of state process goals which are so closely related to the outcome goals that it may be very difficult to judge them separately. For example, the attainment of a process goal like 1 which states that the public schools "shall provide instruction which bears a meaningful relationship to the present and future needs and/or interests of pupils" may be achieved to a certain extent if instruction is directed toward student attainment of all outcome goals. Therefore, to assess attainment of process goal 1 involves assessing the attainment of the outcome goals, at whatever level they are being achieved. In addition, however, such things as checklists, teacher-observation scales to be used by supervisors, lesson plans, questionnaires to be completed by teachers, and staff evaluation of the curriculum and curriculum materials can be used to focus primarily on the attainment of the process goal. Then, the data--on outcome and process goals--should be examined together to identify where needs are. It is important to remember that the process goals are concerned with how the outcome goals are achieved. Although the attainment of a state process goal like 7 that calls for the public schools to provide "resources for education, used with maximum efficiency" can be partially assessed independent of the assessment of the outcome goals (that part dealing with providing resources for education), even here there is a need to relate it to outcome goals since process goal 7 requires assessment as to whether the resources are "used with maximum efficiency." Attainment of state outcome goals is closely related to efficient use of resources. But

you must keep in mind that a school district may be using its resources at a maximum efficiency level and still not be able to attain certain outcome goals at the desired level because of a number of factors, such as insufficient funds for additional needed resources, too high a standard set for attainment of some goals/objectives, problems in attaining other process goals as needed, and so on.

In assessing attainment of state outcome and process goals, it is essential to remember that you are dealing with an interrelated structure. An assessment of needs cannot be made on the basis of outcome goals alone or on the basis of process goals alone. The total picture must be reviewed to identify needs.

Using Data to Identify Needs

Once the data are collected, they need to be organized so that critical needs will readily be seen. On the following pages, Examples D and E from charts of needs-assessment data show how the data might be organized for identifying needs. These examples include, and are extensions of, some of the material found in the Examples A and C from needs-assessment plans that are found earlier in this chapter.

After all the data are organized, they should be reviewed by the professional staff in accordance with the administrative code. The professional staff should examine the data in terms of both the short- and long-range objectives, determining areas of student needs. Needs in relation to the process goals should also be examined at this time. The following procedure might be followed:

- (1) Compare the standards for each objective with the data obtained from the various measures.
- (2) Check in red those objectives for which the levels of proficiency are not currently being met.
- (3) Examine the red-checked objectives and decide, in light of the data, if the proficiency levels set were realistic and/or appropriate. Also examine the indicators used to see if they are appropriate. (A checklist of questions to be asked about

EXAMPLE D FROM CHART OF NEEDS ASSESSMENT DATA FOR 1976-77

GOAL	OBJECTIVES	STANDARDS		SAMPLE	RESULTS OF MEASURES USED		OTHER INDICATORS
		EXPECTANCY LEVELS	MEASURES				
To acquire knowledge, habits and attitudes that promise personal and public health, both physical and mental	Students will <u>not</u> smoke cigarettes.	100% below 8th grade 95% grades 8-9 90% grades 10-11 85% grade 12	Student questionnaire	All	Those admitting to smoking once: 5% in grades 4-7 10% in grades 8-9 25% in grades 10-11 30% in grade 12	Those admitting to regular smoking: 0% in grades 4-7 1% in grades 8-9 5% in grades 10-11 6% in grade 12	Shopkeepers at local shopping centers complain regularly about large numbers of students in grades 8-12 standing around smoking and littering area with cigarette butts.
			Number of students disciplined for smoking on school grounds	All	Disciplined for smoking on school grounds: 0% in grades 4-7 5% in grades 8-9 10% in grades 10-11 10% in grade 12		
			Conference with parents	5% random sample in each grade	Sample of parents indicated: 0% in grades 4-7 smoke regularly 0% in grades 8-9 smoke regularly 1% in grades 10-11 smoke regularly 3% in grade 12 smoke regularly		

Note: If the results of parent conferences appear to be unrealistic, it may be because of possible unwillingness on the part of parents to admit their children smoke, even when they may suspect it to be true.

EXAMPLE E FROM CHART OF NEEDS ASSESSMENT DATA FOR 1976-77

GOALS	OBJECTIVES	STANDARDS		SAMPLE	RESULTS OF MEASURE USED	OTHER INDICATORS
		EXPECTANCY LEVELS	MEASURES			
To acquire basic skills in reading, writing, and mathematics	By fourth grade, students will be able to identify the main idea and supporting details as stated in material read.	85% of all fourth-graders will be able to correctly answer a ratio of 4 out of 5 questions related to this objective on the N.J. Statewide Assessment Reading Test.	<u>N.J. Statewide Assessment Test in Reading</u>	All fourth-grade students - whole population tested	95% of all fourth-graders correctly answered a ratio of 4 out of 5 questions related to this objective on the N.J. Statewide Assessment Reading Test	Records in cumulative folders, including reading test results, records of parent-teacher conferences, samples of teacher made tests results and student work, etc.
		80% of the fourth-graders will be able to correctly respond to all questions on the Informal Reading Inventory.	Informal reading inventory in which students will be asked, after silent reading of a page or less, to read orally the parts that answer specific questions (i.e., what is the main idea, read something that helps support or prove the main idea, etc.).	All fourth-grade students in the following 10 classrooms: <u>School</u> <u>Room Numbers</u> Martin Luther King 101, 104 Thomas Jefferson 211, 212, 214 Jane Adams 106, 109, 111 John F. Kennedy 110, 114 The sample consists of: 1/2 of all fourth-graders; a representative sample in terms of age, sex, race, grades (based on student records), neighborhood characteristics within the school district.	90% of those given an informal reading inventory correctly responded to questions related to this objective	
		80% of the fourth-graders will have <u>no</u> mention of problems on this objective in the teacher's anecdotal records.	Teacher's anecdotal records	Same as sample in 2 above	5% of those in the classes sampled were indicated as having problems in reading related to this objective; in the remaining 95% no problem was indicated related to this objective	

indicators is given earlier in this chapter.) If there is a doubt, you should look at other indicators (if they are available) and/or reexamine the way in which the standards were set. This should be done for the objectives not rechecked to be certain that you are adequately measuring attainment of all objectives.

- (4) Once you have decided that the standards are, indeed, appropriate, put a blue star opposite those objectives which are most critical. The needs related to objectives dealing with the basic communication and computational skills must be planned for if achievement is below the state standards. The other needs may be ranked in terms of priorities set by the district's professional staff, including those needs in the area of the process goals.

Remember, too, that data regarding the statewide assessment program are available from the New Jersey Department of Education. Interpretation Guidelines are available each year that tell you how to interpret the District Level Reports, Building Level Reports, Classroom Level Reports, and other report options.

Notes

1. For Example A, expectancy levels vary by grade. Generally the trend is to increase the expected level by grade. Note, however, that for cigarette smoking the expectancy levels drop with each grade level. The expectancy levels should take account of the fact that more and more students do smoke as they get older regardless of what the school is able to do.

The measures in this example include no standardized tests. Evaluation should include tests where appropriate, but other types of measures such as performance tests, teacher observations, and conferences should be used as well.

The sample column shows that it is not necessary to collect data on every student to evaluate the performance of a district. Of course, as the sample size gets smaller, the amount of random error in the results increases. (See Appendix B-1 for a brief discussion of sample sizes and associated sampling errors.)

2. The levels of proficiency in Examples B and C are in terms of the percentage of students that (a) obtain a given score or higher and (b) perform in a specific way on a specific task or in a specific situation. The measures in these examples consist of standardized tests (external measures), internal measures, and teacher observations.

The sample column shows that it is not necessary to collect data on every student; but where information on every student is available (Example C), it is used. The sample is described in terms of specific classes and student characteristics.

INFORMATION FOR ASSESSMENT AND EVALUATION

CHAPTER 4

PROGRAM IMPROVEMENT/PROGRAM SELECTION

Purpose of This Chapter

The purpose of this chapter is to discuss program improvements through adaptation of current programs, adoption of new programs, and supervision or in-service training that will help you attain your stated goals, objectives, and standards.

Definition of Program Improvement

Your educational program, the curriculum, refers to your district's or school's total plan for learning, the purposes for which are specified by your goals, objectives, and standards. Various curricular programs or areas/courses of study--such as communication skills, mathematics, social studies--make up the total educational program.

A program, as the word is used in this chapter, means more than a set of courses or units of study. It means a plan for learning, arranging for students to be exposed to a certain kind of content (books, films, exhibits, and so on) and certain kinds of experiences and modes of presentation that will help them acquire certain kinds of skills and knowledge. If the needs analysis you have carried out has indicated that your students are not acquiring the skills and knowledge you have specified in your goals and objectives, there may be something wrong with your plan.

Program improvement may involve (1) the adoption or adaptation of an educational program, (2) the process of supervising instructional programs, and/or (3) inservice teacher training to improve the quality of education in your district. This includes making appropriate and necessary revisions to current curricular programs and developing new curricular programs, all in light of the information obtained from the needs assessment.

On the following pages, you will find information on what you can do for program improvement or selection. This information should give you

some ideas about how to improve the curriculum you now have and offer suggestions on how to develop new curricular programs. For additional information, you may wish to consult texts on curriculum development and other resources available from the Association for Supervision and Curriculum Development in Suite 1100, 1701 K Street, N.W., Washington, D.C. 20006.

What Is Required

Your various curricular programs should blend to make an educational program that provides for quality education in your district. More specifically, the New Jersey Administrative Code (NJAC 6:8-3.4) requires that each educational program shall:

1. Be consistent with written goals, objectives, and identified pupil needs.
2. Develop individual talents and interests and serve diverse learning styles to motivate pupil achievement.
3. Provide for continuous learning through an effective articulation between and among the districts and schools.
4. Provide all pupils continuous access to sufficient programs and services of a library/media facility, classroom collection, or both, to support the educational program (curriculum).
5. Provide all pupils guidance and counseling to assist in career and academic planning.
6. Provide a continuum of educational programs and services for all handicapped children, pursuant to law and regulation.
7. Provide bilingual programs for pupils whose dominant language is not English, pursuant to law and regulation.
8. Provide compensatory education programs for pupils, pursuant to law and regulation.
9. Provide all pupils equal educational opportunity, pursuant to law and regulation.
10. Provide career awareness and vocational education, pursuant to law and regulation.

11. Provide educational opportunities for exceptionally gifted and talented pupils.

Should You Change Your Curriculum?

If your curriculum adequately meets the needs of your students in all these respects, should you change it? The answer is no. The T & E legislation does not require all school districts to have the same or even similar programs. Moreover, there is nothing in the legislation that prohibits schools from continuing with existing programs. The only requirement the Administrative Code does make is that all districts select programs that meet the needs of their students in an organized way.

Even if there are apparent weaknesses or inconsistencies in your school's or your district's curriculum, it may not be necessary to develop an entirely new program or to change your program at all. Working within the existing educational program is often possible and more practical. A strong and well-planned program of remediation in reading, for example, might do more to improve students' skills than a new program. Also, educators in areas heavily populated with children whose first language is not English have found that a bilingual approach to curriculum can result in a dramatic improvement in reading.

Two other questions to ask about your curriculum are these: Are your current programs producing the desired student outcomes (Chapter 3 on needs assessment addresses this question.) and have your programs been properly implemented? (See Chapter 5.) There is a considerable difference between a program on paper and one in the classroom. Better implementation of existing programs can be accomplished through supervisor and in-service training.

Balancing Your Educational Program. Your curriculum problem may be an administrative one that can be solved by reorganization. For example, an educational program may be providing too many courses for the teaching staff to handle and for the students to absorb. One way to make such a program more manageable is by combining related subjects into one broad field. This is often done in English by combining spelling, composition,

and literature into a single field of language arts or by combining history and geography into social studies or American history and American literature into American studies. Coordinating programs in this way can increase teaching efficiency and give students what the New Jersey Administrative Code refers to as "a sense of the interdisciplinary nature of knowledge and the interrelatedness of learning."

Seeking Outside Help. If there is a serious weakness in your curriculum or if there is something wrong and you are not certain how to pinpoint it, you may want to seek outside help--from the Educational Improvement Centers (EICs), consultants, or members of your community. Appendix E contains potential sources of information and assistance from outside the immediate community. Many schools conduct in-service training programs to give their faculty and administration new insights, ideas, and techniques about different aspects of education. Another way of enhancing the educational program is through independent study by staff members. A common procedure for approaching curriculum problems is to appoint a committee of faculty, administration, parents, and other interested members of the community to identify problems and suggest ideas, community resources, and guidelines.

What to Look for in a Curriculum

There are a number of things to look for in any curriculum, whether it is a current one you are evaluating or a new one you are thinking of adopting. Here are some of the more important considerations:

Sequence. Every plan for learning involves some kind of sequence or order of learning experiences. As Hilda Taba explains in Curriculum Development: Theory and Practice (New York: Harcourt, Brace, Jovanovich, Inc., 1962), "Certain general principles apply in planning any of these sequences such as the principle of moving from the known to the unknown, from the simple to the complex, from the analysis of concrete experiences to developing generalizations."

Is the content of the school's mathematics program, for example, matched with the appropriate learning experiences to assure a logical

sequence that will enable students to progress from something like simple theorems to complex original problems? And what happens between each learning stage? Is the program designed to give students the opportunity to rest and review what they have learned? Are the steps from one plateau of learning to the next small enough to enable slower students to keep up without boring the more able ones?

Balance and Breadth and Depth. Does the curriculum cover the amount of material it is supposed to cover while at the same time explaining in adequate depth the more important aspects of that material, those principles or underlying truths that the student can apply in other areas of learning? In an English program, a curriculum unit based on Shakespeare's major tragedies, for example, might cover five plays and provide a detailed treatment of certain aspects of those plays that illuminate the nature of tragedy and the drama as well as the structure of poetry. Such a curricular unit should give students not only exposure to some of the world's great drama but also an understanding of theater and poetry that could be applied across all literature.

Appropriateness. A curriculum cannot be viewed without a knowledge of the students who will be exposed to it. Teachers must match curricular programs with students. The tragedies of Shakespeare, while stimulating and effective material for students with high-level reading skills, would probably be totally inappropriate for other students for whom a well-planned unit based on newspapers and magazines would be more constructive and motivating.

Selection of New Courses

It may be that to provide what is needed you will have to add some new courses. This will mean developing your course-content outline and then adopting or adapting a successful program developed by another school district, research and development unit, or publisher, or developing your own materials.

First, you will have to go through a planning stage. In addition to the objectives you have set for your overall program, you must decide upon

another set of specific objectives that relate to the program area or course of study--such as history, English, science--that you need. If you have decided that you need a course to introduce eighth and ninth graders to natural science, for example, what aspects of physics, chemistry, geology, and astronomy do you want them to learn and on what level? Should your course be primarily a descriptive one that will give your students an informed appreciation of science and the scientific method? Or will it involve tasks that require them to do the actual investigative work of physicists and chemists? Will the course call for laboratory as well as classroom work? Will you base the course on one textbook and laboratory projects or use a variety of library resources or possibly have the students devise their own textbooks based on their reports from the field and the laboratory? In other words, your planning for a new course to meet the needs of your students should focus on answering the following questions:

- (1) What are the objectives of the new course and how do these conform with the overall goals and objectives you have for your students?
- (2) What is to be the specific content of the course in light of the objectives?
- (3) Will additional resources be needed if you add this course and, if so, what are they?

Selecting a Successful Program. Today there are hundreds of programs that have been developed by practitioners in school districts across the nation. For most of these programs there is available evidence of effectiveness with students. The EIC in your region can discuss these program options with you, assist you in adapting them to your needs, and provide staff training, if required.

Deciding on Published Material. If you have your course objectives clearly in mind and you decide to buy published material, your most difficult problem will be in choosing the right publisher. There are many publishers in the education field today, and you may have to see several before you find the material that fits your plan. Be sure to find out if the published material has been field tested with students and what the evaluation

results were.

Developing Your Own Material. This is a very different matter.

Curriculum development is a complex subject that has filled large books and cannot be discussed in such a small space without risking oversimplification. The nature of the project will vary, of course, according to the kind of course it is and the objectives it is designed to achieve. Courses developed in the 1950s such as the Chemical Bond Approach Project at Earlham College, the Physical Science Study Committee physics course developed at MIT, and the Harvard Project physics course kept teams of college and high school science teachers busy for as long as six or seven years. On the other hand, some textbooks in English have been written in six months.

Most textbooks are written like any other book. The author begins with an outline or blueprint of what the book will contain and then follows the outline, unit for unit, until the book is completed. If two or more authors are involved, as is often the case, they may review each other's work as they go along. If everything works out as planned, the contents of the material should match the objectives of the course point for point. Before you decide to develop new materials, however, be sure to find out what already exists.

Review and Tryout of New Courses. Whether you are preparing your own course, adopting a program, or using published materials, what you have first is an untried course as far as your district or school is concerned. So the next phase is a careful review and tryout to determine if what you have is what you want and need. This part of curriculum development or implementation can range in scope from a review by committees and a year's tryout by teachers in all schools in the district to review by one or two colleagues and experimentation by one or two classes inside a single school. The purpose of all reviews and tryouts is the same: to determine whether the course does the job it is intended to do and, if not, why not.

Steps In the Program Improvement Process

You cannot begin improving your program until you have identified your needs. Since needs assessment was discussed in Chapter 3, it will not be dealt with here other than to say that when you plan for program improvement each year, you will have to decide which needs you must meet in the given

year and what program areas address them. Once this prerequisite is taken care of, there are three steps to program improvement: analysis of course, development of a plan, and implementation of the plan.

However, before you begin on these steps toward program improvement, you need to answer the question: Did your district/school develop process goals in the goal-development stage? If so, did those goals describe a desirable educational process for developing your program? If the answer is no to either of these questions, you will have to identify your process goals--that is, identify the unique characteristics of the educational process for the program area(s) under consideration. Remember, needs assessment tells you where you are in terms of your goals, objectives, and standards. The difference between (a) where you are and (b) your goals, objectives, and standards defines your needs and their urgency. What is equally essential for program improvement is to know the difference between what your educational practices (or processes) are now and what you want them to be. Knowing this helps you decide how much your program needs to change.

With the needs and the program area(s) that address them identified and the desirable educational practices defined, you are ready to follow the steps toward program improvement. These steps are nicely explained in materials developed by Charles Barthe, Jean Sadenwater, Art Spangenberg and Dottie Wilson of the Educational Improvement Center-Central. The following pages contain some information from their materials.

Analysis of Cause. This step toward program improvement involves the analysis of each existing program that contributes to each need and the identification of possible improvements, including the development of a rationale for those changes. For each program area under consideration, you can develop a worksheet somewhat like the example shown on the following page.

Each member of the staff responsible for instruction in each program area should contribute to the analysis-of-cause process by filling out a worksheet. Everyone filling out the worksheet first refers to the identified needs addressed by the program area and lists in column I the specific program efforts related to each identified need. Each person then decides what changes need to be made and puts these in column II, numbering each suggested

Analysis of Cause Worksheet		
Program Area: <u>Reading</u>		By <u>Jane Doe</u>
I What We Are Doing Now	II What Changes Need To Be Made	III Why
1. Using the <i>Green</i> series as a basal	1. None	1. ———
2. Using supplementary readers in the classroom, all selected by the teachers	2. Enlist the aid of librarians to help students self-select books at appropriate readability levels	2. Develop student interests and motivate reading more
Etc.	Etc.	Etc.

change to correspond to what is listed in column I. For each suggested change, a rationale should be given in column III, numbered to correspond to what is in the first two columns. Once everyone has completed his/her worksheet, the staff can discuss this information and reach consensus on what changes need to be made. This list of proposed changes will become the basis for program-level objectives.

Following is a checklist of some of the things you might change to improve a program:

- (1) Learning styles, such as student interests and achievements
- (2) Teaching styles, such as lecture method, guided-discovery approach, problem solving, and individual/independent study
- (3) Instructional strategies, such as specific programs from publishers or other schools
- (4) Classroom management, such as total-group instruction,

- small-group instruction, and learning centers
- (5) Instructional equipment/material, such as specific media and supplemental materials
 - (6) Organization, such as self-contained classroom, flexible scheduling, and team teaching
 - (7) Physical environment, such as lab areas, areas for large-group instruction, areas for small-group instruction
 - (8) Policy making, such as administrative decisions and board decisions

Remember, this is only a partial list of suggested things you might consider in making program changes. You will, of course, be able to add to this list.

Development of a Plan. Once the analysis-of-cause process is completed and proposed changes are identified as program-level objectives, you are ready to develop a plan for implementing the changes.

First, fill in another worksheet like the one you used in the analysis-of-cause-process but in this one indicate the staff's consensus on what changes need to be made and why. Then have the staff look at all recommended changes in column II and code them as follows:

- (1) Place a US by each statement of recommended change that the group agrees represents a simple change that can be made within the district/school/classroom using existing strategies for supervision of instruction.
- (2) Place a UD by each statement of change that requires in-service teacher training.
- (3) Place an A/A by each statement of recommended change that the group agrees can be made by adopting/adapting an existing program, part of a program, or combination of programs or their parts. (Of course, prior to this, a search of existing programs would have already been made through the EICs, program catalogs, consultants, etc., and possible changes would already have been identified from these sources as well as from a review of the current district/school programs.)

- (4) Place a D by each recommended change that can be made as a result of "in-house" development activity. (Since such activity is usually complex and time consuming, you will probably do this only when nothing else can satisfy your needs.)
- (5) Place a D₂ next to each recommended change that can be made by utilizing outside development resources such as Title IV-C, Compensatory Education Research and Development money. (Remember that this, too, is time consuming and makes additional demands on your developmental process.)

Once the coding above is completed, review column II, the changes needed to be made. If every one is coded, you have identified strategies for making the agreed-upon changes and you are ready to prepare your plan for implementing those changes. If you find, however, that you do not have all the strategies for making essential changes for a specific need, determine whether: (a) it is still possible to make significant gains on your need by making the changes you can make, or (b) you need to put a hold on working on this need for this year and identify what can be worked on in its place, or (c) you will do what is necessary this year to be ready to work on the original need next year.

To make your plan, you might use a format somewhat like the following:

Plan for 19--					
A Program-Level Objectives	B Plan	C Responsible Individuals	D Time Line		E Evaluation
			Start	End	
(Changes to be made)	(Activities)	(Names or Positions)	(Dates)	(Dates)	(Standards)

Transfer your change statements (column II of your Analysis of Cause Worksheet) to column A of the plan since these are your program-level objectives. In column B and next to each change, list the activities to be carried out to make the change. Column C should list the person or position (e.g., principal of each school) responsible for each change, while column D shows the time span required to carry out the activities for each change. Respond to column E, the evaluation, by answering the following questions:

- (1) How will you know if you are successful and to what degree you have been successful?
- (2) How will you know what is right and what went wrong with the process?

Chapter 5 deals specifically with program evaluation which you should include in your plan for program improvement.

Implementation of the Plan. Now you are ready to implement your plan. Carry out the plans you outlined and monitor progress through on-going/formative evaluation. You may find that changes will be needed. Make the necessary changes but keep your colleagues in your school, the district office, and the county superintendent's office informed about them. You will find them and the EICs helpful in implementing your plan.

INFORMATION FOR ASSESSMENT AND EVALUATION

CHAPTER 5
PROGRAM EVALUATION



Purpose of This Chapter

The purpose of this chapter is to explain program evaluation. The chapter will define evaluation and related terms, indicate the legal requirements, and discuss aspects of designing an evaluation including reference to what kinds of data are collected and how they are collected.

Definition of Evaluation

Evaluation is a systematic process to determine the value and success of programs, projects, and materials in relation to the achievement of goals, objectives, and standards set by the district (NJAC 6:8-1.1). It includes the collection of both qualitative and quantitative data plus the making of value judgments about that data.

Evaluation is an essential part of the administration of any educational activity--from a single classroom lesson to a massive curriculum development project--for without it, you cannot be faithful to the goal of quality education.

The purpose of any evaluation is to provide information to decision makers so that decisions about what to do next can be based on facts rather than guesses or hunches.

Every program, project, or activity in a school system has some specific goal or goals which should have been clear and explicit before the activity began. But often the goals of an activity are unclear or unknown. In the evaluation process, the first task is to define the goals of the program to be evaluated and state them as specific objectives. Goals are the things we continually aspire to; objectives are the accomplishments we actually expect to reach. Chapters 1 and 2 focused on goal setting and objective setting respectively.

Once you know the goals and objectives of a program, you must then define the procedures and activities that are, or will be, utilized to achieve the objectives. ("Program Implementation/Selection," Chapter 4, addressed this.) When you know the objectives and procedures of the program, you then design techniques for gathering information about reaching those objectives and recording the procedures. This is your evaluation design and the data-gathering techniques you select will mean locating and developing measuring instruments such as tests, pupil record forms, and so forth. Material in Chapter 3, "Needs Assessment," provided information regarding internal and external measures. Of course, it is best and easiest to plan your program evaluation when you plan program implementation/selection because then you can collect data about procedures and the reaching of objectives as part of the program itself. At intervals during the project and at its end, you summarize the data, draw conclusions about the program's effectiveness, and make decisions about what changes should be made in the future.

Legal Requirements for Evaluation

According to the Administrative Code, evaluation procedures are to be developed and implemented to provide for the continuous and comprehensive review of pupil progress toward district and school goals and program objectives. The evaluation is to be conducted by teaching staff members under the direction of the chief school administrator and should include consultation with parents or guardians of pupils. Further, the results of the district's evaluation plan are to be reported annually to the district board of education at a public meeting (NJAC 6:8-3.7).

Some Key Concepts

When people hear the word "evaluation," negative and unpleasant images often come to mind. The kind of evaluation that is most threatening and unpleasant is the "pass-fail" or "good-bad" kind. The technical name for this is summative evaluation, a clear-cut investigation of whether something is good or bad, effective or ineffective, worthwhile or worthless,

workable or unworkable. In education, summative evaluation helps determine whether materials are to be used or rejected and programs continued, modified, or dropped.

In formative evaluation, the purpose of which is to improve a program as you evaluate it, you implement changes based on your findings as the program progresses rather than waiting until the end to make a decision. Persons trained in experimental methodology are often unhappy with this process; the experimentalist wants to take a detached view of a whole project without "contaminating" it with ongoing changes. In the real educational world, however, programs change as they progress (whether we wish it or not), and the task of evaluation is to keep track of these changes and analyze their effects rather than complain about confounding variables.

In most modern notions of educational management, curriculum development, and evaluation, "objectives" is a key concept. Most human enterprises--certainly all educational enterprises--have objectives. But in many instances, the objectives of an educational program are unclear or unknown. Identification of objectives should be part of all planning and decision-making processes. Even if a program has no explicit objectives, the evaluator of the program must endeavor to identify them. Objectives are those milestones that we do, in fact, reach on the path to the goals. An objective is an event we wish to have occur.

Defining What Is To Be Evaluated

Before the current national concern about educational evaluation, most educational plans and proposals consisted of lists of procedures and practices. The traditional curriculum guide devotes little space to pupil-behavior objectives but discusses at length the procedures and materials that will be given or applied to the student. It explains what teachers will do to students but not what should happen to students as a result.

Modern evaluation methods include human-performance changes as program objectives; but you must still specify the procedures, practices, techniques, and materials you will use to reach those objectives. When you evaluate a program, you are determining whether the practices you employed resulted in

achieving the desired objectives. If you find that a program has reached its objectives but you have neglected to analyze and record the procedures and techniques that were used, you have failed in your evaluation. On the other hand, if a program fails to meet its objectives, it may be that the proposed techniques never were employed, in which case you failed again because you still do not know if these techniques are effective. Evaluation, then, must collect data about both objectives and procedures.

A procedure should be described clearly and specifically enough so that the evaluator can be certain it is taking place. To say that a program uses the "inquiry approach" or "an individualized technique" does not really indicate clearly what is happening in the program, since these terms can be construed in hundreds of ways. As with performance objectives, all terms used in the description of a procedure should be defined in their operational context. Moreover, you should know the criteria for deciding whether or not the particular procedure (activity, approach, technique) specified in your program is, in fact, being followed--especially if the procedure is novel or subject to many interpretations.

Some or all of the following characteristics are useful in defining any procedure:

- People performing the technique: teachers, specialists, aides, or others; their special skills and backgrounds as they relate to the program
- People affected by the program: pupils, teachers, parents, or anyone else on the receiving end of the program; special attributes of the people affected; basis for deciding which persons are selected or served
- Practices: what the persons operating the program actually do and what the people affected actually do
- Materials and facilities: what equipment, supplies, institutional materials, rooms, buildings, etc., are employed in the program
- Expenditures: what the cost is of each program activity; what costs are officially charged to the program; what charges are paid from other sources in the budget; what the expenditures are

for program time, space, depreciation, etc.

- **Schedule:** the periods in which certain activities take place; the projected deadline dates for completing activities and achieving objectives

The objectives of a program are the payoff of its activities; they are variously known as the effects, outputs, products, behavioral changes, and impact. By any name, they are the reason for being of the program, the desired outcomes against which the value of the program is assessed. If the objectives are well-stated, the evaluation of their achievement is straightforward--either by direct measure of the behavior stipulated in the objective or indirect inference based on the measurement of some related variables. Most affective and attitudinal objectives are in this latter class; we make inferences about student motivation by recording changes in dropout rate, rate of participation in co-curricular activities, or by a questionnaire. Although some people believe that questionnaires constitute a direct measure of attitude, they are really an indirect measure of mental states, and this should be recognized in the design or interpretation of self-report questionnaires or opinionnaires. The evaluation design should stipulate what direct or indirect measures will be used to evaluate the outcomes of the program.

An essential attribute of all well-formed objectives is that you can know with a degree of certainty whether or not they have been achieved. The evaluator will be able to determine this by observation, test, or some other measure. It is easiest to evaluate objectives for which there are instruments with proven utility. Even so, in stipulating criteria for determining whether objectives have been achieved, it is permissible to use appeals to expert judges or panels of experienced educators. But the more subjectivity involved in determining whether an objective has been reached, the more likely you are to start controversies about your evaluative conclusions. The criteria should be sufficiently explicit that even persons who doubt the efficacy of the program will agree when they are confronted with the evidence that the objective was met.

Designing the Evaluation Plan

An educational program rarely consists of one procedure with one outcome. Therefore, to carry out the evaluation of a program, you should stipulate what differences there are within it. Evaluation of the program should answer these questions:

- Outcome objectives--Will the people affected by the program exhibit the same performance changes as a result of the program; can we group the people affected by the program according to specified expectations?
- Procedural factors--Will all the people working in the program do the same things to the same people? Are there important differences between classes or sites? Are materials and supplies used in the same way throughout the program?
- Phases--Is the program really a sequence of smaller subprograms, each with its own objectives and procedures? What are the logical parts into which the program can be divided?

Although they may not all appear under the evaluation section of a plan or proposal, the following elements should be involved in an evaluation design:

1. Statement of program objectives in clear, rationally assessable terms
2. Definition and description of procedures, including intraprogram differentiation
3. Schedule showing patterns of activity, objective deadlines, and times
4. Procedures for assessing whether each objective has been met
5. Record forms for keeping track of people performing, people receiving, and procedural techniques
6. A schedule for conducting the evaluation activities, including testing days and report schedule
7. A budget for the evaluation activities--usually 5 percent

of the program budget, but more if the effectiveness of the program is highly important or highly uncertain

The Experimental Method: One of Several Approaches to Program Evaluation

The experimental method is one approach to program evaluation. It is, perhaps, the most sophisticated and complex approach. This particular approach has been selected for further discussion since the basic concepts of the experimental method are relevant to many other approaches to program evaluation. Needless to say, each district must devise and implement program evaluations which are practical, realistic, and responsive to the district's own information needs. It is assumed that districts' program evaluations will continuously improve as practical experience is gained.

In the experimental method, the outcomes of the program are called the dependent variables. The forces and factors designed to produce them are the independent variables. Independent variables are the people, materials, facilities, and other components of the program which we manage in order to produce the desired outcomes. It is incorrect to think of a program as a single independent variable; the evaluation design should list the set of functionally discrete procedures involved as a series of independent variables and construct a model or paradigm which relates each of the set of independent variables to each of the outcomes or dependent variables.

In a sense, the purpose of an evaluative study is to test whether or not some presumably causal relationship exists between the independent and dependent variables. Thus, the design of an evaluation is based on one of several hypotheses which are stated in this form:

Some variation (or change in condition or magnitude) in each independent variable will produce some variation in the dependent variable, in some particular direction (higher or lower), under certain contingent conditions. (For example: A pupil-teacher ratio of 20-1 will result in significantly higher language arts performance than a pupil-teacher ratio of 30-1 for disadvantaged first graders.)

With the experimental approach, you should state as many hypotheses as

there are independent variables or combinations of variables you wish to evaluate, in terms of their effectiveness. Here is another example:

Showing the Captain Science TV series with introductions and summaries by teachers will produce significantly greater science achievement among fifth graders than showing the series without such introductions and summaries.

The nature of experimental, statistical research is such that you cannot directly prove that your hypotheses are true for false but only reject or not reject the "null" versions of these hypotheses. To illustrate:

Research Hypothesis: "Homogeneous grouping of students by achievement areas on the Iowa Test of Basic Skills will result in a significantly greater gain in reading skills than heterogeneous grouping for elementary school students."

Null Hypotheses: "There will be no significant difference in improvement of reading skills for heterogeneously and homogeneously grouped elementary school students."

In your evaluation study, you will test the null assertion; as a result of your study, you will be able either to reject the null hypothesis with some specific level of certainty and infer that your research hypothesis is true or not reject the null hypothesis and be unable to conclude that your research hypothesis is true.

Indicating Confidence in Your Measuring Instruments. The word significant is used frequently in evaluation, particularly when the experimental approach is used. "Significant difference" or "significant outcome" does not mean a large difference or an important outcome. In statistics, "significant" in this context means that the result is very probably a reflection of a real difference or a real change in the population being measured rather than a result of chance. When statisticians say that scores are "significantly different at the .01 level or $p < .01$," they mean that if the same investigation had been conducted 100 times, they would have found

no difference in scores no more than one time out of the hundred.

Significance, then, is a measure of the confidence you have in your measuring procedures. In interpreting evaluation data, you must decide whether your outcomes are important enough to be the basis for decisions about change. Using statistical "significance" helps you to do this.*

Showing Change by a Pretest-Posttest Design. Evaluation studies of successful programs should show change in the dependent variable, and in order to show change, at least two observations, measures, or tests are required. The pretest-posttest approach is generally used in which the same measure is applied before and after the procedures are employed, and the difference in scores is the measure of improvement. In the absence of a pretest (and there are occasionally good reasons for not giving them), some historical facts about the people affected by the program will do, such as previous test scores or behaviors. The historical data should be accurate and sound and, as far as possible, based on the same measures employed in the posttest.

It is true that even with the pre- and posttest model of program evaluation you may wonder how much more students learned by participating in a program than they would have without it. To help deal with this question, G. Kasten Tallmadge and Christine T. Wood prepared the User's Guide: ESEA Title I Evaluation and Reporting System published in 1976 by RMC Research Corporation for the Office of Education, U.S. Department of Health, Education and Welfare. Their basic evaluation model is to determine program effect by identifying the difference between observed posttest performance and expected no-program performance. Appendix B-3 contains some information derived from this material which you may find useful. Keep in mind, however, that like any other evaluation model, this one has its limitations, probably the most serious one being that results on standardized norm-referenced measures are essential to using the model. Certainly, this restricts your use of the model to student achievement data. Yet you cannot ignore other

*Ideally, a school should consult a statistician for help with these decisions.

measures in your evaluation plan. Just as with the needs assessment, nontest measures are important to obtain the full picture. Your evaluation plan should contain both test and nontest measures and include those measures in the standards you set. Further, you should concern yourself with other things about the program, such as cost effectiveness, impact on the district/school organization and staff, articulation with other programs, and incidental and unintended effects on students.

Comparing and Measuring Groups. Almost every evaluation study attempts to compare something with something else. Either you are comparing the effectiveness of two or more sets of procedures or you are comparing the effectiveness of one set of procedures or "chance." For this reason, many evaluation designs, particularly those using the experimental approach, compare the changes in two or more groups--groups which are alike except for the independent variables. Like grouping can generally be achieved in two ways: either by matching the two groups with respect to relevant contingent variables (age, previous training, etc.) or by random assignment to the various categories of independent variables. Each group that receives some subset of the experimental procedures is a Treatment Group or Experimental Group, and the one that receives no treatment and is presumably affected by ordinary or "chance" influences is called a Control Group.

The procedures followed in making such comparisons should be reliable. Differences in measures and observations on the same population should reflect differences in the population, not unevenness of the measurement procedure itself. Judges and scorers should treat the same information the same way; the distribution of test scores for a given population should change systematically and understandably rather than randomly or erratically. The absence of these requirements means that the evaluation procedure was unreliable, and the information gathered is useless.

Validity refers to the degree to which measures, tests, and the whole experimental design itself measure what they are supposed to measure, and not something else. Most tests measure some variable or other, but is it

the one you are interested in? If not, how closely does it correlate with the variable you are interested in?

Even the most vigorous evaluation designs cannot anticipate all the influences that might affect the outcome of the program. When you examine the results, certain contingencies or contingent variables may appear to be important, often variables that you did not incorporate into the design, such as class size, socioeconomic status, or teacher attitudes. These contingencies should be built into your conclusions as limitations on the generalizability of the findings. A good evaluation design is limited by contingencies.

Intervening variables are entities or processes that happen during the experimental procedure that may alter or influence the outcome. Usually these are mental processes that one cannot fully know or understand but that may be important contingencies in the study, especially if they are not present in all your treatment groups. Generally, any variable that was not part of the design scheme--anything from a peculiarity of attitude to a bizarre current event--whose precise effect cannot be determined but that affects the outcome of the program is a confounding variable. You cannot assess its importance, and, usually, it prevents you from assessing the effectiveness of the independent variables as well. In a thorough evaluation, some care should be given to determining the good effects of the program that were not anticipated and the bad effects that were not predicted. For example, a math program in a school may stimulate parents to continue their own education, an unanticipated good effect which should be mentioned; but the same math program may have caused high anxiety in students, or negatively affected some math skills previously acquired, a bad effect which should also be mentioned.

Some Final Thoughts

An evaluation should help everyone associated with it, and the information it produces should be communicated to all the people who might profit from it. The results of an evaluation can, of course, affect the program in important ways. If a program is controversial, for example, the more facts

and valid inferences there are about it, the higher will be the level of debate.

Evaluation activities should not be secret. Persons affected by the evaluation should be informed about it and, indeed, involved in its planning and execution. And if the evaluation yields negative or embarrassing information, it should be discussed with the people affected by it before it is published or disseminated.

Despite its scientific and statistical trappings, good program evaluation can be a way of humanizing educational management by stimulating open communication and cooperation in the process of finding out what is true, what is good, and how education can be continually improved.

INFORMATION FOR ASSESSMENT AND EVALUATION

PART III

APPENDICES



The Evaluation Improvement Program

APPENDIX A-1 - Matrix of Instruments Matched
to New Jersey State Goals

The purpose of this matrix is to indicate which of the twelve New Jersey State Educational Outcome Goals are covered by a particular measurement instrument listed in APPENDIX A-2. The goals are listed on the left-hand side of the page; test code numbers are listed across the top. Refer to APPENDIX A-2, Matrix of Test Descriptions, to obtain the test title that corresponds to each code number. For example, the test referred to by the test code number 10 on page 101 is Assessment of Career Development listed on page 115. A test is designated as appropriate to a specific goal if the publisher has indicated that it was designed specifically to assess attainment of that goal, or if it contains subtests or items that can be used for that purpose. The instruments included in the Matrix of Test Descriptions were selected to provide ample representation across all outcome goals.

New Jersey State Educational Goals

Test Code Number

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
1. To acquire basic skills in obtaining information, solving problems, thinking critically and communicating effectively.		✓				✓								✓	✓	✓		✓		✓	
2. To acquire a stock of basic information concerning the principles of the physical, biological and social sciences, the historical record of human achievements and failures and current social issues.					✓		✓														
3. To become an effective and responsible contributor to the decision-making processes of the political and other institutions of the community, state, country and world.					✓																
4. To acquire the knowledge, skills and understanding that permit him/her to play a satisfying and responsible role as both producer and consumer.										✓											✓
5. To acquire job entry level skills and, also, to acquire knowledge necessary for further education.		✓								✓											✓
6. To acquire the understanding of and the ability to form responsible relations with a wide range of other people, including but not limited to those with social and cultural characteristics different from his/her own.				✓									✓								
7. To acquire the capacities for playing satisfying and responsible roles in family life.				✓																	
8. To acquire the knowledge, habits and attitudes that promote personal and public health, both physical and mental.	✓			✓													✓				✓
9. To acquire the ability and the desire to express himself/herself creatively in one or more of the arts, and to appreciate the aesthetic expressions of other people.											✓								✓		
10. To acquire an understanding of ethical principles and values and the ability to apply them to his/her own life.																					
11. To develop an understanding of his/her own worth, abilities, potentialities and limitations.		✓	✓	✓				✓	✓	✓	✓										✓
12. To learn to enjoy the process of learning and to acquire the skills necessary for a lifetime of continuous learning and adaptation to change.								✓	✓			✓	✓								✓

New Jersey State Educational Goals

Test Code Number

	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42
1. To acquire basic skills in obtaining information, solving problems, thinking critically and communicating effectively.					✓				✓		✓	✓		✓	✓	✓	✓	✓	✓	✓	✓
2. To acquire a stock of basic information concerning the principles of the physical, biological and social sciences, the historical record of human achievements and failures and current social issues.											✓	✓	✓					✓			
3. To become an effective and responsible contributor to the decision-making processes of the political and other institutions of the community, state, country and world.																					
4. To acquire the knowledge, skills and understanding that permit him/her to play a satisfying and responsible role as both producer and consumer.		✓	✓	✓	✓					✓		✓				✓					
5. To acquire job entry level skills and, also, to acquire knowledge necessary for further education.	✓	✓	✓	✓	✓					✓		✓	✓								
6. To acquire the understanding of and the ability to form responsible relations with a wide range of other people, including but not limited to those with social and cultural characteristics different from his/her own.						✓	✓														
7. To acquire the capacities for playing satisfying and responsible roles in family life.							✓														
8. To acquire the knowledge, habits and attitudes that promote personal and public health, both physical and mental.							✓														
9. To acquire the ability and the desire to express himself/herself creatively in one or more of the arts, and to appreciate the aesthetic expressions of other people.									✓												
10. To acquire an understanding of ethical principles and values and the ability to apply them to his/her own life.								✓													
11. To develop an understanding of his/her own worth, abilities, potentialities and limitations.	✓			✓	✓			✓													
12. To learn to enjoy the process of learning and to acquire the skills necessary for a lifetime of continuous learning and adaptation to change.									✓												

	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63
1. To acquire basic skills in obtaining information, solving problems, thinking critically and communicating effectively.	✓			✓	✓		✓	✓	✓					✓		✓	✓	✓		✓	✓
2. To acquire a stock of basic information concerning the principles of the physical, biological and social sciences, the historical record of human achievements and failures and current social issues.		✓	✓							✓									✓		
3. To become an effective and responsible contributor to the decision-making processes of the political and other institutions of the community, state, country and world.			✓																		
4. To acquire the knowledge, skills and understanding that permit him/her to play a satisfying and responsible role as both producer and consumer.																					
5. To acquire job entry level skills and, also, to acquire knowledge necessary for further education.																					
6. To acquire the understanding of and the ability to form responsible relations with a wide range of other people, including but not limited to those with social and cultural characteristics different from his/her own.											✓	✓	✓								
7. To acquire the capacities for playing satisfying and responsible roles in family life.																					
8. To acquire the knowledge, habits and attitudes that promote personal and public health, both physical and mental.																					
9. To acquire the ability and the desire to express himself/herself creatively in one or more of the arts, and to appreciate the aesthetic expressions of other people.																					
10. To acquire an understanding of ethical principles and values and the ability to apply them to his/her own life.															✓						
11. To develop an understanding of his/her own worth, abilities, potentialities and limitations.						✓															
12. To learn to enjoy the process of learning and to acquire the skills necessary for a lifetime of continuous learning and adaptation to change.																					

New Jersey State Educational Goals

Test Code Number

	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84
1. To acquire basic skills in obtaining information, solving problems, thinking critically and communicating effectively.	✓		✓			✓	✓						✓						✓	✓	
2. To acquire a stock of basic information concerning the principles of the physical, biological and social sciences, the historical record of human achievements and failures and current social issues.			✓	✓	✓		✓	✓													
3. To become an effective and responsible contributor to the decision-making processes of the political and other institutions of the community, state, country and world.			✓	✓																	
4. To acquire the knowledge, skills and understanding that permit him/her to play a satisfying and responsible role as both producer and consumer.			✓									✓									
5. To acquire job entry level skills and, also, to acquire knowledge necessary for further education.			✓							✓											✓
6. To acquire the understanding of and the ability to form responsible relations with a wide range of other people, including but not limited to those with social and cultural characteristics different from his/her own.			✓		✓									✓							
7. To acquire the capacities for playing satisfying and responsible roles in family life.			✓									✓		✓			✓	✓			
8. To acquire the knowledge, habits and attitudes that promote personal and public health, both physical and mental.			✓						✓			✓		✓	✓	✓	✓				
9. To acquire the ability and the desire to express himself/herself creatively in one or more of the arts, and to appreciate the aesthetic expressions of other people.			✓								✓										
10. To acquire an understanding of ethical principles and values and the ability to apply them to his/her own life.			✓	✓					✓					✓							
11. To develop an understanding of his/her own worth, abilities, potentialities and limitations.			✓						✓	✓				✓							✓
12. To learn to enjoy the process of learning and to acquire the skills necessary for a lifetime of continuous learning and adaptation to change.	✓	✓							✓		✓			✓							

New Jersey State Educational Goals

Test Code Number

	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100	101	102	103	104	105
1. To acquire basic skills in obtaining information, solving problems, thinking critically and communicating effectively.		✓					✓			✓				✓	✓						
2. To acquire a stock of basic information concerning the principles of the physical, biological and social sciences, the historical record of human achievements and failures and current social issues.					✓										✓		✓				✓
3. To become an effective and responsible contributor to the decision-making processes of the political and other institutions of the community, state, country and world.					✓										✓		✓				
4. To acquire the knowledge, skills and understanding that permit him/her to play a satisfying and responsible role as both producer and consumer.			✓			✓						✓									
5. To acquire job entry level skills and, also, to acquire knowledge necessary for further education.			✓			✓						✓									
6. To acquire the understanding of and the ability to form responsible relations with a wide range of other people, including but not limited to those with social and cultural characteristics different from his/her own.	✓										✓				✓	✓			✓	✓	
7. To acquire the capacities for playing satisfying and responsible roles in family life.	✓														✓	✓					
8. To acquire the knowledge, habits and attitudes that promote personal and public health, both physical and mental.								✓	✓						✓				✓		✓
9. To acquire the ability and the desire to express himself/herself creatively in one or more of the arts, and to appreciate the aesthetic expressions of other people.				✓											✓			✓			
10. To acquire an understanding of ethical principles and values and the ability to apply them to his/her own life.								✓			✓										
11. To develop an understanding of his/her own worth, abilities, potentialities and limitations.	✓		✓			✓		✓	✓			✓	✓		✓	✓			✓		
12. To learn to enjoy the process of learning and to acquire the skills necessary for a lifetime of continuous learning and adaptation to change.									✓						✓	✓			✓		

New Jersey State Educational Goals

Test Code Number

	106	107	108	109	110	111	112	113	114	115	116	117	118	119	120	121	122	123	124	125	126
1. To acquire basic skills in obtaining information, solving problems, thinking critically and communicating effectively.	✓	✓			✓	✓	✓					✓					✓	✓	✓		
2. To acquire a stock of basic information concerning the principles of the physical, biological and social sciences, the historical record of human achievements and failures and current social issues.																					
3. To become an effective and responsible contributor to the decision-making processes of the political and other institutions of the community, state, country and world.																					
4. To acquire the knowledge, skills and understanding that permit him/her to play a satisfying and responsible role as both producer and consumer.											✓			✓	✓						
5. To acquire job entry level skills and, also, to acquire knowledge necessary for further education.														✓	✓						
6. To acquire the understanding of and the ability to form responsible relations with a wide range of other people, including but not limited to those with social and cultural characteristics different from his/her own.				✓						✓											
7. To acquire the capacities for playing satisfying and responsible roles in family life.										✓										✓	
8. To acquire the knowledge, habits and attitudes that promote personal and public health, both physical and mental.										✓											
9. To acquire the ability and the desire to express himself/herself creatively in one or more of the arts, and to appreciate the aesthetic expressions of other people.			✓					✓					✓			✓					✓
10. To acquire an understanding of ethical principles and values and the ability to apply them to his/her own life.																					
11. To develop an understanding of his/her own worth, abilities, potentialities and limitations.										✓					✓						
12. To learn to enjoy the process of learning and to acquire the skills necessary for a lifetime of continuous learning and adaptation to change.									✓												

New Jersey State Educational Goals

Test Code Number

	127	128	129	130	131	132	133	134	135	136	137	138	139	140	141	142	143	144	145	146	147
1. To acquire basic skills in obtaining information, solving problems, thinking critically and communicating effectively.			✓	✓				✓					✓		✓	✓					✓
2. To acquire a stock of basic information concerning the principles of the physical, biological and social sciences, the historical record of human achievements and failures and current social issues.			✓									✓				✓			✓	✓	
3. To become an effective and responsible contributor to the decision-making processes of the political and other institutions of the community, state, country and world.														✓		✓			✓		
4. To acquire the knowledge, skills and understanding that permit him/her to play a satisfying and responsible role as both producer and consumer.						✓						✓	✓	✓							
5. To acquire job entry level skills and, also, to acquire knowledge necessary for further education.						✓															
6. To acquire the understanding of and the ability to form responsible relations with a wide range of other people, including but not limited to those with social and cultural characteristics different from his/her own.							✓														
7. To acquire the capacities for playing satisfying and responsible roles in family life.					✓		✓														
8. To acquire the knowledge, habits and attitudes that promote personal and public health, both physical and mental.					✓		✓										✓	✓			
9. To acquire the ability and the desire to express himself/herself creatively in one or more of the arts, and to appreciate the aesthetic expressions of other people.	✓	✓						✓	✓												
10. To acquire an understanding of ethical principles and values and the ability to apply them to his/her own life.																					
11. To develop an understanding of his/her own worth, abilities, potentialities and limitations.						✓	✓				✓										
12. To learn to enjoy the process of learning and to acquire the skills necessary for a lifetime of continuous learning and adaptation to change.							✓														

New Jersey State Educational Goals

Test Code Number

	148	149	150	151	152	153	154	155	156	157	158	159	160	161	162	163	164	165	166	167	168
1. To acquire basic skills in obtaining information, solving problems, thinking critically and communicating effectively.	✓											✓	✓							✓	✓
2. To acquire a stock of basic information concerning the principles of the physical, biological and social sciences, the historical record of human achievements and failures and current social issues.																					
3. To become an effective and responsible contributor to the decision-making processes of the political and other institutions of the community, state, country and world.																					
4. To acquire the knowledge, skills and understanding that permit him/her to play a satisfying and responsible role as both producer and consumer.	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓										✓	
5. To acquire job entry level skills and, also, to acquire knowledge necessary for further education.	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓										✓	
6. To acquire the understanding of and the ability to form responsible relations with a wide range of other people, including but not limited to those with social and cultural characteristics different from his/her own.										✓					✓		✓	✓			
7. To acquire the capacities for playing satisfying and responsible roles in family life.														✓	✓		✓				
8. To acquire the knowledge, habits and attitudes that promote personal and public health, both physical and mental.															✓	✓	✓		✓		
9. To acquire the ability and the desire to express himself/herself creatively in one or more of the arts, and to appreciate the aesthetic expressions of other people.																					
10. To acquire an understanding of ethical principles and values and the ability to apply them to his/her own life.																		✓			
11. To develop an understanding of his/her own worth, abilities, potentialities and limitations.									✓	✓				✓		✓		✓	✓	✓	
12. To learn to enjoy the process of learning and to acquire the skills necessary for a lifetime of continuous learning and adaptation to change.											✓					✓					

New Jersey State Educational Goals

Test Code Number

	169	170	171	172	173	174	175	176	177	178	179	180	181	182	183	184	185	186	187	188	189
1. To acquire basic skills in obtaining information, solving problems, thinking critically and communicating effectively.	✓	✓			✓			✓	✓	✓	✓	✓	✓	✓	✓	✓					
2. To acquire a stock of basic information concerning the principles of the physical, biological and social sciences, the historical record of human achievements and failures and current social issues.				✓										✓			✓	✓			
3. To become an effective and responsible contributor to the decision-making processes of the political and other institutions of the community, state, country and world.				✓										✓				✓			
4. To acquire the knowledge, skills and understanding that permit him/her to play a satisfying and responsible role as both producer and consumer.						✓								✓				✓			
5. To acquire job entry level skills and, also, to acquire knowledge necessary for further education.						✓															
6. To acquire the understanding of and the ability to form responsible relations with a wide range of other people, including but not limited to those with social and cultural characteristics different from his/her own.					✓														✓	✓	
7. To acquire the capacities for playing satisfying and responsible roles in family life.																					
8. To acquire the knowledge, habits and attitudes that promote personal and public health, both physical and mental.																		✓	✓		
9. To acquire the ability and the desire to express himself/herself creatively in one or more of the arts, and to appreciate the aesthetic expressions of other people.							✓							✓	✓						
10. To acquire an understanding of ethical principles and values and the ability to apply them to his/her own life.						✓															
11. To develop an understanding of his/her own worth, abilities, potentialities and limitations.		✓			✓	✓													✓	✓	
12. To learn to enjoy the process of learning and to acquire the skills necessary for a lifetime of continuous learning and adaptation to change.					✓		✓														✓

New Jersey State Educational Goals

Test Code Number

	190	191	192	193	194	195	196	197	198	199	200	201	202	203	204	205	206	207	208	209	210
1. To acquire basic skills in obtaining information, solving problems, thinking critically and communicating effectively.														✓	✓			✓			✓
2. To acquire a stock of basic information concerning the principles of the physical, biological and social sciences, the historical record of human achievements and failures and current social issues.	✓													✓	✓						✓
3. To become an effective and responsible contributor to the decision-making processes of the political and other institutions of the community, state, country and world.	✓													✓	✓						
4. To acquire the knowledge, skills and understanding that permit him/her to play a satisfying and responsible role as both producer and consumer.										✓	✓			✓	✓				✓		
5. To acquire job entry level skills and, also, to acquire knowledge necessary for further education.										✓	✓								✓		
6. To acquire the understanding of and the ability to form responsible relations with a wide range of other people, including but not limited to those with social and cultural characteristics different from his/her own.		✓	✓									✓					✓			✓	
7. To acquire the capacities for playing satisfying and responsible roles in family life.		✓										✓									
8. To acquire the knowledge, habits and attitudes that promote personal and public health, both physical and mental.		✓		✓	✓	✓		✓					✓						✓		
9. To acquire the ability and the desire to express himself/herself creatively in one or more of the arts, and to appreciate the aesthetic expressions of other people.							✓									✓					
10. To acquire an understanding of ethical principles and values and the ability to apply them to his/her own life.																					
11. To develop an understanding of his/her own worth, abilities, potentialities and limitations.								✓	✓	✓	✓	✓	✓								
12. To learn to enjoy the process of learning and to acquire the skills necessary for a lifetime of continuous learning and adaptation to change.			✓	✓	✓	✓			✓			✓									

New Jersey State Educational Goals

Test Code Number

	211	212	213	214	215	216	217	218	219	220	221	222	223	224	225	226	227	228	229	230	231
1. To acquire basic skills in obtaining information, solving problems, thinking critically and communicating effectively.	✓	✓	✓	✓	✓	✓			✓												✓
2. To acquire a stock of basic information concerning the principles of the physical, biological and social sciences, the historical record of human achievements and failures and current social issues.	✓																		9		✓
3. To become an effective and responsible contributor to the decision-making processes of the political and other institutions of the community, state, country and world.																		✓			
4. To acquire the knowledge, skills and understanding that permit him/her to play a satisfying and responsible role as both producer and consumer.							✓											✓		✓	
5. To acquire job entry level skills and, also, to acquire knowledge necessary for further education.							✓														
6. To acquire the understanding of and the ability to form responsible relations with a wide range of other people, including but not limited to those with social and cultural characteristics different from his/her own.										✓									✓		
7. To acquire the capacities for playing satisfying and responsible roles in family life.										✓									✓		
8. To acquire the knowledge, habits and attitudes that promote personal and public health, both physical and mental.							✓			✓					✓	✓			✓		
9. To acquire the ability and the desire to express himself/herself creatively in one or more of the arts, and to appreciate the aesthetic expressions of other people.														✓			✓				
10. To acquire an understanding of ethical principles and values and the ability to apply them to his/her own life.										✓	✓										
11. To develop an understanding of his/her own worth, abilities, potentialities and limitations.						✓			✓						✓	✓					
12. To learn to enjoy the process of learning and to acquire the skills necessary for a lifetime of continuous learning and adaptation to change.							✓					✓	✓								

Test Code Number

[illegible]

APPENDIX A-2 - Matrix of Test Descriptions

This matrix is intended to provide the user with a brief overview of a variety of tests. It indicates which of several characteristics considered important in the test selection process apply to each instrument. When a characteristic is not marked, it means that it is not pertinent to the particular test. In two instances (population and administration time), specific information is provided. The publisher of the test is designated by a number. Refer to the list in Appendix A-4 for the publisher's name and complete address. Abstracts providing a description of the instruments and their technical properties have been prepared for many but not all of the tests. The final column lists the page number on which an abstract appears if it is included in this notebook.

Measures	Population	Type of Administration		Alternate Forms Available	Scoring		Availability of Scoring Services	Administration Time	Materials Available							Scores Available										Publisher	Page of Test Description (if available)
		Group	Individual		Machine	Hand			Specimen Set	Reusable Test Booklet	One-time Test Booklet	Administrator's Manual	Interpretive Material	Practice Material	Reliability Information Available	Validity Information Available	Standard Scores	Percentiles	Grade Equivalents	Age Equivalents	Stanines	Profiles/Scales	Raw Scores	Subscores	Not Indicated		
1. AAPHER Cooperative Health Education Test	Gr. 4-6	X			X	X		40 mins.	X	X		X													X	1	147
2. ACT Career Planning Program	16 Yrs.+	X			X	X	X	147 mins.	X			X	X							X	X					2	147
3. About Me	Gr. 4-6	X				X		Un-timed			X				X										X	60	-
4. Adjustment Inventory: Student Form	Gr. 9-16	X	X		X	X		Un-timed	X	X	X	X	X		X	X		X					X	X		12	147
5. American Political Behavior Achievement Test	Gr. 9-12	X				X		90-100 mins.	X		X	X			X								X			25	148
6. Analysis of Readiness Skills: Reading and Mathematics	K-Gr. 1	X	X			X		30-40 mins.	X		X	X	X		X			X					X			27	148
7. Anderson-Fisk Chemistry Test	Gr. 10-13	X		E,F	X	X	X	40 mins.	X	X		X	X				X	X			X		X			45	-
8. Animal Crackers	Pre K-Gr. 1	X	X		X	X	X	30-60 mins.	X	X	X	X		X	X	X		X				X	X			9	149
9. Arlin-Hills Attitude Survey	K-Gr. 3	X				X		10 mins.			X	X	X		X	X		X	X				X	X		5	149
10. Assessment of Career Development	Gr. 8-11	X			X		X	125 mins.	X	X		X	X		X	X						X	X	X		27	149
11. Attitude Questionnaire	Gr. 7-8	X	X			X		Un-timed			X		X												X	65	150
12. Attitude Toward Classroom Atmosphere	Gr. 1-6	X				X		15-25 mins.			X	X			X	X									X	17	150

Measures	Population	Type of Administration		Alternate Forms Available	Scoring		Availability of Scoring Services	Administration Time	Materials Available						Reliability Information Available	Validity Information Available	Scores Available								Publisher	Page of Test Description (if available)	
		Group	Individual		Machine	Hand			Specimen Set	Reusable Test Booklet	One-time Test Booklet	Administrator's Manual	Interpretive Material	Practice Material			Standard Scores	Percentiles	Grade Equivalents	Age Equivalents	Stanines	Profiles/Scales	Raw Scores	Subscores			Not Indicated
13. Barclay Classroom Climate Inventory	Gr. 2-6	X			X		X	75 mins.	X	X		X	X		X	X						X				18	151
14. Basic Skills in Arithmetic	Gr. 6-12	X		A,B		X		40 mins.	X		X	X	X		X			X					X			55	151
15. Blyth Second-Year Algebra Test	Gr. 9-12	X			X	X	X	40 mins.	X	X		X	X				X	X			X		X			45	-
16. Boehm Test of Basic Concepts	K-Gr.2	X	X	A,B		X		30 mins.	X		X	X	X		X	X		X	X					X		45	151
17. Bristol Social Adjustment Guides	5-15 yrs.	X						Un-timed			X	X	X		X	X		X			X	X			16	-	
18. Brown-Carlson Listening Comprehension Test	Gr. 9-13+	X		Am, Bm	X	X	X	45-50 mins.	X			X	X		X		X	X				X	X	X		45	152
19. Bryant-Schwan Design Test	Pre K-Gr. 12		X			X		10 mins.					X		X	X								X	10	152	
20. California Achievement Tests: 1970 Edition	Gr. 1.5-12	X		A,B	X	X	X	120-153 mins.	X	X	X	X	X		X	X		X	X		X	X	X	X		9	153
Level 1	Gr. 1.5-2	X		A,B	X	X	X	114 mins.	X	X	X	X	X		X	X		X	X		X	X	X	X		9	-
Level 2	Gr. 2-4	X		A,B	X	X	X	123 mins.	X	X	X	X	X		X	X		X	X		X	X	X	X		9	-
Level 3	Gr. 4-6	X		A,B	X	X	X	152 mins.	X	X	X	X	X		X	X		X	X		X	X	X	X		9	-
Level 4	Gr. 6-9	X		A,B	X	X	X	144 mins.	X	X	X	X	X		X	X		X	X		X	X	X	X		9	-

Measures	Population	Type of Administration		Alternate Forms Available	Scoring		Availability of Scoring Services	Administration Time	Materials Available							Reliability Information Available	Validity Information Available	Scores Available							Publisher	Page of Test Description (if available)	
		Group	Individual		Machine	Hand			Specimen Set	Reusable Test Booklet	One-time Test Booklet	Administrator's Manual	Interpretive Material	Practice Material	Standard Scores			Percentiles	Grade Equivalents	Age Equivalents	Stanines	Profiles/Scales	Raw Scores	Subscores			Not Indicated
Level 5	Gr. 9-12	X		A,B	X	X	X	153 mins.	X	X	X	X	X		X	X		X	X		X	X	X	X		9	-
21. California Life Goals Evaluation Schedules	16 yrs.+	X	X	A,B,C,D-M,D-S	X	X		45 mins.	X	X		X			X	X		X			X	X	X		63	153	
22. California Occupational Preference Survey	Gr. 9-16+	X	X		X	X		30-40 mins.	X	X	X	X	X		X	X		X			X	X	X		16	154	
23. Career Awareness Inventory	Gr. 4-12	X			X	X	X	60-90 mins.	X	X		X			X	X					X		X	X		54	154
24. Career Development Inventory: Form I	Gr. 7-14	X			X	X		30 mins.		X		X			X	X		X				X	X		58	154	
25. Career Maturity Inventory	Gr. 6-12	X			X	X	X	65 mins.	X	X		X	X		X	X		X	X			X	X	X		9	155
26. Career Planning Program	Gr. 8-11	X			X		X	Not Given	X	X		X	X		X	X					X	X	X	X		27	155
27. Chapin Social Insight Test	13 yrs.	X				X		Un-timed	X	X			X			X								X	12	-	
28. Child Behavior Rating Scale	3-9 yrs.	X	X			X		Un-timed	X	X		X			X	X						X	X		63	156	
29. Children's Self-Conception Test	Pre K-Gr. 3	X	X	I,II		X		120-150 mins.		X		X	X											X	60	-	
30. CIRCUS	Pre K-K	X	X		X	X	X	Var-ies	X		X	X	X	X	X	X							X	X		1	156
31. Comprehensive Career Assessment Scales	Gr. 3-12	X	X			X		20-30 mins.			X	X	X		X	X					X		X		33	157	

Measures	Population	Type of Administration		Alternate Forms Available	Scoring		Availability of Scoring Services	Administration Time	Materials Available							Reliability Information Available	Validity Information Available	Scores Available							Publisher	Page of Test Description (if available)	
		Group	Indiv. Qual		Machine	Hand			Specimen Set	Reusable Test Booklet	One-time Test Booklet	Administrator's Manual	Interpretive Material	Practice Material	Standard Scores			Percentiles	Grade Equivalents	Age Equivalents	Stanines	Profiles/Scales	Raw Scores	Subscores			Not Indicated
32. Comprehensive Tests of Basic Skills: Forms S and T	K-Gr. 12	X		S,T	X	X	X	145-268 mins.	X	X	X	X	X	X	X	X		X	X		X	X	X	X		9	157
Level A	K-Gr.1.3	X		S,T	X	X	X	144 mins.	X	X	X	X	X	X	X	X		X	X		X	X	X	X		9	157
Level B	K.6-Gr.1.9	X		S,T	X	X	X	154 mins.	X	X	X	X	X	X	X	X		X	X		X	X	X	X		9	157
Level C	Gr.1.6-2.9	X		S,T	X	X	X	240 mins.	X	X	X	X	X	X	X	X		X	X		X	X	X	X		9	157
Level 1	Gr.2.5-4.9	X		S,T	X	X	X	268 mins.	X	X	X	X	X	X	X	X		X	X		X	X	X	X		9	158
Level 2	Gr.4.5-6.9	X		S,T	X		X	263 mins.	X	X	X	X	X	X	X	X		X	X		X	X	X	X		9	158
Level 3	Gr.6.5-8.9	X		S,T	X		X	262 mins.	X	X	X	X	X	X	X	X		X	X		X	X	X	X		9	158
Level 4	Gr.8.5-12.9	X		S,T	X		X	255 mins.	X	X	X	X	X	X	X	X		X	X		X	X	X	X		9	158
33. Content Evaluation Series	Gr.7-12	X			X	X	X	40-160 mins.	X	X		X	X		X	X	X	X			X		X	X		27	158
34. Agribusiness Achievement Test	Gr.9-12	X			X	X	X	160 mins.	X	X	X	X	X		X	X	X	X			X		X	X		27	158
35. Language Arts Tests	Gr.7-12	X			X	X	X	120 mins.	X	X	X	X	X		X	X	X	X			X		X	X		27	158
36. Mathematics Test	Gr.7-9	X			X	X	X	40 mins.	X	X	X	X	X		X	X	X	X			X		X	X		27	159
37. Modern Economics Test	Gr.10-12	X			X	X	X	45 mins.	X	X	X	X	X		X	X	X	X			X		X	X		27	159

Measures	Population	Type of Administration		Alternate Forms Available	Scoring		Availability of Scoring Services	Administration Time	Materials Available						Reliability Information Available	Validity Information Available	Scores Available								Publisher	Page of Test Description (if available)	
		Group	Individual		Machine	Hand			Specimen Set	Reusable Test Booklet	One-time Test Booklet	Administrator's Manual	Interpretive Material	Practice Material			Standard Scores	Percentiles	Grade Equivalents	Age Equivalents	Stanines	Profiles/Scales	Raw Scores	Subscores			Not Indicated
38. Modern Geometry Test	Gr.10-12	X			X	X	X	45 mins.	X	X	X	X	X		X	X	X	X			X		X	X		27	159
39. Science Tests	Gr. 8-9	X			X	X	X	40 mins.	X	X	X	X	X		X	X	X	X			X		X	X		27	150
40. Cooperative Achievement Tests	Gr.9-14	X		Var-ies	X	X	X	Var-ies	X	X		X	X		X	X		X					X			1	160
41. Cooperative English Tests		X		A, B,C	X	X	X	40 mins.	X	X	X	X	X		X	X		X					X	X		1	160
Level 1	Gr.12-14	X		A, B,C	X	X	X	40 mins.	X	X	X	X	X		X	X		X					X	X		1	-
Level 2	Gr.9-12	X		A, B,C	X	X	X	40 mins.	X	X	X	X	X		X	X		X					X	X		1	-
42. Cooperative Literature Tests	Gr.9-12	X		A,B		X		40 mins.	X	X		X	X		X	X							X			1	160
43. Cooperative Mathematics Tests	Gr.7-14	X		Var-ies	X	X	X	Var-ies	X	X		X	X		X	X		X					X			1	160
Arithmetic	Gr.7-9	X		A, B,C	X	X	X	40 mins.	X	X		X	X		X	X		X					X			1	-
Structure of the Number System	Gr.7-8	X		A,B	X	X	X	40 mins.	X	X		X	X		X	X		X					X			1	-
Algebra I	Gr.8-9	X		A,B	X	X	X	40 mins.	X	X		X	X		X	X		X					X			1	-
Algebra II	Gr.9-12	X		A,B	X	X	X	40 mins.	X	X		X	X		X	X		X					X			1	-
Geometry	Gr.9-12	X		A,B	X	X	X	80 mins.	X	X		X	X		X	X		X					X			1	-

Measures	Population	Type of Administration		Alternate Forms Available	Scoring		Availability of Scoring Services	Administration Time	Materials Available						Reliability Information Available	Validity Information Available	Scores Available								Publisher	Page of Test Description (if available)	
		Group	Individual		Machine	Hand			Specimen Set	Reusable Test Booklet	One-time Test Booklet	Administrator's Manual	Interpretive Material	Practice Material			Standard Scores	Percentiles	Grade Equivalents	Age Equivalents	Stanines	Profiles/Scales	Raw Scores	Subscores			Not Indicated
Trigonometry	Gr. 9-14	X		A, B	X	X	X	40 mins.	X	X		X	X		X	X		X					X			1	-
Algebra III	Gr. 9-14	X		A, B	X	X	X	40 mins.	X	X		X	X		X	X		X					X			1	-
Analytic Geometry	Gr. 9-14	X		A, B	X	X	X	40 mins.	X	X		X	X		X	X		X					X			1	-
Calculus	Gr. 9-14	X		A, B	X	X	X	80 mins.	X	X		X	X		X	X		X					X			1	-
44. Cooperative Science Tests	Gr. 9-12	X		A, B	X	X	X	80 mins.	X	X		X			X	X		X					X			1	160
Advanced General Science Test	Gr. 8-9	X		A, B	X	X	X	80 mins.	X	X		X			X	X		X					X			1	160
Biology Test	Gr. 9-12	X		A, B	X	X	X	80 mins.	X	X		X			X	X		X					X			1	161
Chemistry Test	Gr. 10-12	X		A, B	X	X	X	80 mins.	X	X		X			X	X		X					X			1	161
General Science Test	Gr. 7-9	X		A, B	X	X	X	80 mins.	X	X		X			X	X		X					X			1	161
Physics Test	Gr. 9-12	X		A, B	X	X	X	80 mins.	X	X		X			X	X		X					X			1	161
45. Cooperative Social Studies Tests	Gr. 7-12	X		A, B		X	X	40 mins.	X	X		X			X	X		X					X			1	162
American History Test	Gr. 7-8	X		A, B		X	X	40 mins.	X	X		X			X	X		X					X			1	162
American Government Test	Gr. 10-12	X		A, B		X	X	40 mins.	X	X		X			X	X		X					X			1	162

Measures	Population	Type of Administration		Alternate Forms Available	Scoring			Administration Time	Materials Available						Reliability Information Available	Validity Information Available	Scores Available								Publisher	Page of Test Description (if available)
		Group	Individual		Machine	Hand	Availability of Scoring Services		Specimen Set	Reusable Test Booklet	One-time Test Booklet	Administrator's Manual	Interpretive Material	Practice Material			Standard Scores	Percentiles	Grade Equivalents	Age Equivalents	Stanines	Profiles/Scales	Raw Scores	Subscores		
Civics Test	Gr.8-9	X		A,B		X	X	40 mins.	X	X		X		X	X		X					X			1	162
Modern European History Test	Gr.10-12	X		A,B		X	X	40 mins.	X	X		X		X	X		X					X			1	163
Problems of Democracy Test	Gr.10-12	X		A,B		X	X	40 mins.	X	X		X		X	X		X					X			1	163
World History Test	Gr.10-12	X		A,B		X	X	40 mins.	X	X		X		X	X		X					X			1	163
46. Cooperative Preschool Inventory	Ages 3-6		X			X		15 mins.	X		X	X	X	X		X						X	X		1	163
47. Cooperative Primary Tests	Gr.1-3	X		A,B	X	X		210 mins.	X		X	X	X	X	X	X	X					X	X		1	164
Pilot	Gr.1-3	X		A,B	X	X		10 mins.	X		X	X	X	X	X	X	X					X			1	164
Reading	Gr.1-3	X		A,B	X	X		35 mins.	X		X	X	X	X	X	X	X					X	X		1	164
Listening	Gr.1-3	X		A,B	X	X		35 mins.	X		X	X	X	X	X	X	X					X	X		1	164
Word Analysis	Gr.1-3	X		A,B	X	X		40 mins.	X		X	X	X	X	X	X	X					X	X		1	165
Writing Skills	Gr.1-3	X		A,B	X	X		40 mins.	X		X	X	X	X	X	X	X					X	X		1	165
Mathematics	Gr.1-3	X		A,B	X	X		50 mins.	X		X	X	X	X	X	X	X					X	X		1	164
48. Coopersmith Self-Esteem Inventory	9 yrs.+	X		A,B		X		Un-timed			X		X		X							X	X		56	165

Measures	Population	Type of Administration		Alternate Forms Available	Scoring		Availability of Scoring Services	Administration Time	Materials Available						Reliability Information Available		Validity Information Available		Scores Available								Publisher	Page of Test Description (if available)
		Group	Individual		Machine	Hand			Specimen Set	Reusable Test Booklet	One-time Test Booklet	Administrator's Manual	Interpretive Material	Practice Material	Standard Scores	Percentiles	Grade Equivalents	Age Equivalents	Stanines	Profiles/Scales	Raw Scores	Subscores	Not Indicated					
49. Cornell Class-Reasoning Test	Gr. 4-12	X					Un-timed			X											X			14	-			
50. Cornell Conditional Reasoning Test	Gr. 4-12	X					Un-timed			X											X			14				
51. Cornell Critical Thinking Test	Gr. 7-12	X	X		X		50 mins.			X	X	X		X	X		X	X			X			14	-			
52. Cray American History Test: Revised Edition	Gr. 10-13	X		E,F	X	X	40 mins.	X	X	X	X	X					X			X		X		45	-			
53. Cross-Cultural Attitude Inventory	3-12 yrs.	X	X			X	20-30 mins.			X	X				X						X			15	165			
54. Cultural Attitude Inventories	Gr. 4+	X				X	20-30 mins.	X	X		X			X	X						X			33	166			
55. Cultural Attitude Scales	5-14 yrs.	X				X	15-20 mins.	X		X	X	X		X	X		X				X			33	166			
56. Davis Reading Test	Gr. 8-13	X		1,2	X	X	40 mins.	X	X		X			X	X		X				X	X		45	166			
57. Defining Issues Test	Gr. 9+	X	X		X	X	50-60 mins.		X		X	X		X	X	X					X			51	167			
58. Diagnostic Mathematics Inventory	Gr.1.5-8.5	X			X		Un-timed	X	X		X	X	X								X	X		9	167			
Level A/Red	Gr.1.5-2.5	X			X		Un-timed	X	X		X	X	X								X	X		9	167			
Level B/Green	Gr.2.5-3.5	X			X		Un-timed	X	X		X	X	X								X	X		9	167			
Level C/Blue	Gr.3.5-4.5	X			X		Un-timed	X	X		X	X	X								X	X		9	167			

Measures	Population	Type of Administration		Alternate Forms Available	Scoring		Availability of Scoring Services	Administration Time	Materials Available						Reliability Information Available	Validity Information Available	Scores Available								Publisher	Page of Test Description (if available)	
		Group	Individual		Machine	Hand			Specimen Set	Reusable Test Booklet	One-time Test Booklet	Administrator's Manual	Interpretive Material	Practice Material			Standard Scores	Percentiles	Grade Equivalents	Age Equivalents	Stanines	Profiles/Scales	Raw Scores	Subscores			Not Indicated
Level D/Orange	Gr. 4.5-5.5	X			X		X	Un-timed	X	X		X	X	X								X	X			9	168
Level E/Aqua	Gr. 5.5-6.5	X			X		X	Un-timed	X	X		X	X	X								X	X			9	168
Level F/Purple	Gr. 6.5-7.5	X			X		X	Un-timed	X	X		X	X	X								X	X			9	168
Level G/Brown	Gr. 7.5-8.5	X			X		X	Un-timed	X	X		X	X	X								X	X			9	168
59. Diagnostic Reading Scales: Revised Edition	Gr. 1-8		X			X		Un-timed	X		X	X	X		X	X						X	X			9	168
60. Diagnostic Tests and Self-Helps in Arithmetic	Gr. 3-12	X	X					Un-timed	X		X	X			X	X						X	X			9	168
61. Dunning-Abeles Physics Test	Gr. 10-13	X	X	E, F	X	X	X	40 mins.	X	X	X	X			X	X	X	X			X		X			45	169
62. Durrell Analysis of Reading Difficulty: New Edition	Gr. 1-6		X			X		30-45 mins.	X		X	X											X	X		45	169
63. Durrell Listening-Reading Series	Gr. 1-9	X		D, E, F	X	X	X	140-170 mins.	X	X	X	X			X	X		X	X	X	X		X	X		45	170
Primary Level	Gr. 1-2	X		D, E, F	X	X	X	140 mins.	X	X	X	X			X	X		X	X	X	X		X	X		45	170
Intermediate Level	Gr. 3-6	X		D, E, F	X	X	X	170 mins.	X	X	X	X			X	X		X	X	X	X		X	X		45	170
Advanced Level	Gr. 7-9	X		D, E, F	X	X	X	160 mins.	X	X	X	X			X	X		X	X	X	X		X	X		45	170

Measures	Population	Type of Administration		Alternate Forms Available	Scoring		Availability of Scoring Services	Administration Time	Materials Available						Reliability Information Available	Validity Information Available	Scores Available							Publisher	Page of Test Description (if available)		
		Group	Individual		Machine	Hand			Specimen Set	Reusable Test Booklet	One-time Test Booklet	Administrator's Manual	Interpretive Material	Practice Material			Standard Scores	Percentiles	Grade Equivalents	Age Equivalents	Stanines	Profiles/Scales	Raw Scores			Subscores	Not Indicated
64. Durrell-Sullivan Reading Capacity and Achievement Test	Gr.2.5-6.9	X		A,B		X		55-95 mins.	X		X	X			X				X	X			X	X		45	170
Primary Level	Gr.2.5-4.5	X		A,B		X		55-65 mins.	X		X	X			X				X	X			X	X		45	-
Intermediate Level	Gr.3-6	X		A,B		X		75-96 mins.	X		X	X			X				X	X			X	X		45	-
65. Education Apperception Test	Pre K-Gr. 6		X			X		Un-timed	X			X													X	63	170
66. Educational Goal Attainment Test	Gr.7-12	X			X	X	X	100 mins.		X		X			X								X	X		44	171
67. Civics	Gr.7-12	X			X	X	X	100 mins.		X		X			X								X	X		44	171
68. Human Relations	Gr.7-12	X			X	X	X	100 mins.		X		X			X								X	X		44	171
69. English Language	Gr.7-12	X			X	X	X	100 mins.		X		X			X								X	X		44	171
70. Reasoning	Gr.7-12	X			X	X	X	100 mins.		X		X			X								X	X		44	172
71. General Knowledge	Gr.7-12	X			X	X	X	100 mins.		X		X			X								X	X		44	172
72. Self	Gr.7-12	X			X	X	X	100 mins.		X		X			X								X	X		44	172
73. Careers	Gr.7-12	X			X	X	X	100 mins.		X		X			X								X	X		44	172

Measures	Population	Type of Administration		Alternate Forms Available	Scoring		Availability of Scoring Services	Administration Time	Materials Available						Reliability Information Available	Validity Information Available	Scores Available								Publisher	Page of Test Description (if available)
		Group	Individual		Machine	Hand			Specimen Set	Reusable Test Booklet	One-time Test Booklet	Administrator's Manual	Interpretive Material	Practice Material			Standard Scores	Percentiles	Grade Equivalents	Age Equivalents	Stanines	Profiles/Scales	Raw Scores	Subscores		
74. Arts and Leisure	Gr.7-12	X			X	X	X	100 mins.		X		X		X							X	X		44	172	
75. Life Skills	Gr.7-12	X			X	X	X	100 mins.		X		X		X							X	X		44	172	
76. Latin America	Gr.7-12	X			X	X	X	100 mins.		X		X		X							X	X		44	173	
77. Ego-Ideal and Conscience Development Test	12-18 yrs.	X			X	X		60 mins.	X	X		X		X	X						X	X		37	173	
78. Emporia Elementary Health Test	Gr.6-8	X		IA, IB, IIA, IIB		X		30 mins.	X		X	X		X		X					X			7	-	
79. Emporia High School Health Test	Gr.9-12	X		A,B		X		40 mins.	X		X	X		X		X					X			7	-	
80. Family Adjustment Test	3-15 yrs.	X	X			X		35-40 mins.		X		X		X	X	X					X	X		48	173	
81. Family Relations Test	3-15 yrs.					X		40 mins.				X		X									X	38	-	
82. Gates-MacGinitie Reading Tests	Gr. 1-12				X	X	X	7-45 mins.	X			X	X	X		X	X	X			X			59	-	
Primary A	Gr. 1	X		1,2	X	X	X	40 mins.	X			X	X		X		X	X			X			59	-	
Primary B	Gr. 2	X		1,2	X	X	X	40 mins.	X			X	X		X		X	X			X			59	-	
Primary C	Gr. 3	X		1,2	X	X	X	40 mins.	X			X	X		X		X	X			X			59	-	

Measures	Population	Type of Administration		Alternate Forms Available	Scoring		Availability of Scoring Services	Administration Time	Materials Available						Reliability Information Available		Validity Information Available		Scores Available								Publisher	Page of Test Description (if available)
		Group	Individual		Machine	Hand			Specimen Set	Reusable Test Booklet	One-time Test Booklet	Administrator's Manual	Interpretive Material	Practice Material	Reliability Information Available	Validity Information Available	Standard Scores	Percentiles	Grade Equivalents	Age Equivalents	Stanines	Profiles/Scales	Raw Scores	Subscores	Not Indicated			
Primary CS	Gr.2-3	X		1, 2, 3	X	X	X	7 mins.	X			X	X		X		X	X	X				X			59	-	
Survey D	Gr.4-6	X		1, 2, 3	X	X	X	45 mins.	X			X	X		X		X	X	X				X			59	-	
Survey E	Gr.7-9	X		1, 2, 3	X	X	X	45 mins.	X			X	X		X		X	X	X				X			59	-	
Survey F	Gr.10-12	X		1, 2	X	X	X	45 mins.	X			X	X		X		X	X	X				X			59	-	
83. Gates-MacGinitie Reading Tests: Readiness Skills	K-Gr.1	X					X	Not Given	X			X			X		X	X	X				X			59	-	
84. Geist Picture Interest Inventory	Gr.8+	X	X	**		X	X	Un-timed	X		X	X			X	X					X	X	X			63	174	
85. Getting Along	Gr.7-9	X		A,B		X		35-45 mins.		X		X	X		X	X					X	X	X			60	174	
86. Gilmore Oral Reading Test: New Edition	Gr.1-8		X	C,D		X		15-20 mins.	X	X		X			X	X		X	X		X		X			45	174	
87. Gordon Occupational Checklist	Gr.9-12	X				X		20-25 mins.	X		X	X			X	X							X	X		45	175	
88. Graves Design Judgment Test	Gr.7+	X				X		20-30 mins.	X	X		X			X	X		X					X			45	175	
89. Grieg Social Studies Test	Gr.6-8	X			X	X		40 mins.	X	X		X			X	X		X					X			54	175	
90. Hall Occupational Orientation Inventory: Second Edition	Gr.7+	X	X		X	X	X	30 mins.	X	X		X	X		X	X					X	X	X			54	176	
91. Harrison-Stroud Reading Readiness Profiles	K-Gr.1	X				X		80 mins.	X		X	X				X		X					X	X		27	176	

** = Male, Female

Measures	Population	Type of Administration		Alternate Forms Available	Scoring		Availability of Scoring Services	Administration Time	Materials Available						Reliability Information Available	Validity Information Available	Scores Available								Publisher	Page of Test Description (if available)
		Group	Individual		Machine	Hand			Specimen Set	Reusable Test Booklet	One-time Test Booklet	Administrator's Manual	Interpretive Material	Practice Material			Standard Scores	Percentiles	Grade Equivalents	Age Equivalents	Stanines	Profiles/Scales	Raw Scores	Subscores		
92. Hartman Value Profile	12 yrs.	X	X			X		10-12 mins.	X		X	X	X		X						X	X	X		49	176
93. How I See Myself Scale	Gr. 3-12	X				X		Not Given	X		X	X			X	X	X						X		23	177
94. Howell Geometry Test	Gr. 9-12	X		A, B	X	X	X	40 mins.	X	X		X			X		X			X		X			45	177
95. Human Relations Inventory	Gr. 9+	X				X		Un-timed	X		X	X			X	X						X			48	-
96. Individual Career Exploration: Experimental Edition	Gr. 8-14	X	X			X		120 mins.	X		X	X									X	X			54	177
97. Inferred Self-Concept Scale: Experimental Form	Gr. 1-6	X		A-2, B-2				Un-timed	X		X	X			X	X	X		X				X		52	178
98. Instructional Objectives Exchange Objectives-Based Tests	K-12	X	X	A, B		X		Var-ies	X		X	X	X		X	X						X			29	173
Language Arts	K-6	X	X	A, B		X		Var-ies	X		X	X	X		X	X			X				X		29	178
Mechanics and Usage	K-6	X	X	A, B		X		Var-ies	X		X	X	X		X	X			X				X		29	178
Word Forms and Syntax	K-6	X	X	A, B		X		Var-ies	X		X	X	X		X	X			X				X		29	179
Composition, Library and Literary Skills	K-6	X	X	A, B		X		Var-ies	X		X	X	X		X	X			X				X		29	179
Mathematics	K-9	X	X	A, B		X		Var-ies	X		X	X	X		X	X			X				X		29	179
Sets and Numbers	K-6	X	X	A, B		X		Var-ies	X		X	X	X		X	X			X				X		20	179

Measures	Population	Type of Administration		Alternate Forms Available	Scoring		Availability of Scoring Services	Administration Time	Materials Available						Reliability Information Available	Validity Information Available	Scores Available							Publisher	Page of Test Description (if available)		
		Group	Individual		Machine	Hand			Specimen Set	Reusable Test Booklet	One-time Test Booklet	Administrator's Manual	Interpretive Material	Practice Material			Standard Scores	Percentiles	Grade Equivalents	Age Equivalents	Stanines	Profiles/Scales	Raw Scores			Subscores	Not Indicated
Operations and Properties	K-6	X	X	A,B		X		Var- ies	X		X	X	X		X	X							X			29	179
Numerations and Relations	K-6	X	X	A,B		X		Var- ies	X		X	X	X		X	X							X			29	179
Measurement	K-6	X	X	A,B		X		Var- ies	X		X	X	X		X	X							X			29	179
Geometry	K-6	X	X	A,B		X		Var- ies	X		X	X	X		X	X							X			29	179
Elements, Symbolism, and Measurement	Gr.7-9	X	X	A,B		X		Var- ies	X		X	X	X		X	X							X			29	179
Geometry, Operations and Relations	Gr.7-9	X	X	A,B		X		Var- ies	X		X	X	X		X	X							X			29	179
Reading	K-6	X	X	A,B		X		Var- ies	X		X	X	X		X	X							X			29	179
Word Attack Skills	K-6	X	X	A,B		X		Var- ies	X		X	X	X		X	X							X			29	179
Comprehension Skills	K-6	X	X	A,B		X		Var- ies	X		X	X	X		X	X							X			29	180
Social Studies	Gr.10-12	X	X	A,B		X		Var- ies	X		X	X	X		X	X							X			29	180
American Government	Gr.10-12	X	X	A,B		X		Var- ies	X		X	X	X		X	X							X			29	180
99. Instructional Objectives Exchange Objectives Collection	K-Gr.12	X				X		Un- timed	X														X			29	180
100. Measures of Self-Concept	K-Gr.12	X				X		Un- timed	X														X			29	180

Measures	Population	Type of Administration		Alternate Forms Available	Scoring			Administration Time	Materials Available						Reliability Information Available	Validity Information Available	Scores Available								Publisher	Page of Test Description (if available)	
		Group	Individual		Machine	Hand	Availability of Scoring Services		Specimen Set	Reusable Test Booklet	One-time Test Booklet	Administrator's Manual	Interpretive Material	Practice Material			Standard Scores	Percentiles	Grade Equivalents	Age Equivalents	Stanines	Profiles/Scales	Raw Scores	Subscores			Not Indicated
101. Objectives Collection in American Government	Gr. 9-12	X				X		Un-timed	X														X			29	180
102. Objectives Collection in Art	Gr. 4-6	X				X		Un-timed	X														X			29	181
103. Objectives Collection in Attitude Toward School	K-Gr. 6	X				X		Un-timed	X					X	X								X			29	181
104. Objectives Collection in Attitudes Related to Tolerance	Gr. 9-12	X				X		Un-timed	X						X								X			29	181
105. Objectives Collection in Health	K-Gr. 6	X				X		Un-timed	X														X			29	181
106. Objectives Collection in Judgment: Analyzing Fallacies and Weaknesses in Arguments	Gr. 7-12	X				X		Un-timed	X														X			29	181
107. Objectives Collection in Judgment: Deductive Logic and Assumption Recognition	Gr. 7-12	X				X		Un-timed	X														X			29	182
108. Objectives Collection in Music	K-Gr. 6	X				X		Un-timed	X														X			29	182
109. Inter-Person Perception Test	6 yrs. +	X		AA, AC		X		15 mins.	X		X	X		X									X			37	183
110. Iowa Silent Reading Tests: 1973 Edition	Gr. 6-16	X		E, F	X	X		56-91 mins.	X	X		X	X		X	X	X	X			X	X	X			45	183
Level 1	Gr. 6-9	X		E, F	X	X		91 mins.	X	X		X	X		X	X	X	X			X	X	X			45	-

Measures	Population	Type of Administration		Alternate Forms Available	Scoring		Availability of Scoring Services	Administration Time	Materials Available						Reliability Information Available	Validity Information Available	Scores Available								Publisher	Page of Test Description (if available)		
		Group	Individual		Machine	Hand			Specimen Set	Reusable Test Booklet	One-time Test Booklet	Administrator's Manual	Interpretive Material	Practice Material			Standard Scores	Percentiles	Grade Equivalents	Age Equivalents	Stanines	Profiles/Scales	Raw Scores	Subscores			Not Indicated	
Level 2	Gr. 9-14	X		E, F	X	X		86 mins.	X	X		X	X		X	X	X	X		X	X	X			45	-		
Level 3	Gr. 11-16	X		E, F	X	X		56 mins.	X	X		X	X		X	X	X	X		X	X	X			45	-		
111. Iowa Tests of Basic Skills: Primary Battery	Gr. 1.7-3.5	X		5, 6	X	X	X	392 mins.	X	X	X	X	X	X	X	X		X	X		X	X	X	X		27	184	
112. Iowa Tests of Basic Skills: Levels Edition	Gr. 3-8	X		5, 6	X	X	X	279 mins.	X	X	X	X	X	X	X	X	X	X	X		X	X	X			27	-	
113. Iowa Tests of Music Literacy	Gr. 4-12	X			X	X		90 mins.	X	X		X			X	X	X	X					X	X		7	184	
114. JIM Scale	Gr. 7-12	X				X		30 mins.				X	X		X	X		X						X		60	184	
115. Jesness Behavior Checklist	13 yrs.+	X			X	X		10-20 mins.	X	X	X	X			X	X		X						X	X		12	185
116. Junior High School Test of Economics	Gr. 7-9	X			X	X		40 mins.	X	X		X	X		X	X		X						X		30	185	
117. KeyMath Diagnostic Arithmetic Test	K-Gr. 7		X			X		30 mins.	X	X		X	X	X				X			X	X	X			3	186	
118. Knauber Art Ability Test	Gr. 7+	X				X		Not Given				X	X												X		32	-
119. Knowledge of Occupations Test	Gr. 9-12	X			X	X		40 mins.	X	X		X			X	X		X					X	X		47	186	
120. Kuder General Interest Survey	Gr. 6+	X			X	X		30-40 mins.	X	X	X	X	X		X	X		X				X	X	X		55	186	
121. Kvalvasser Music Talent Test	Gr. 4-16	X		A, B		X		10 mins.	X		X	X					X							X		36	187	

Measures	Population	Type of Administration		Alternate Forms Available	Scoring			Administration Time	Materials Available						Reliability Information Available	Validity Information Available	Scores Available								Publisher	Page of Test Description (if available)
		Group	Individual		Machine	Hand	Availability of Scoring Services		Specimen Set	Reusable Test Booklet	One-time Test Booklet	Administrator's Manual	Interpretive Material	Practice Material			Standard Scores	Percentiles	Grade Equivalents	Age Equivalents	Stanines	Profiles/Scales	Raw Scores	Subscores		
122. Lankton First-Year Algebra Test: Revised Edition	Gr.8-12	X		E,F	X	X	X	40 mins.	X	X		X		X	X		X		X		X				45	187
123. Lee-Clark Reading Readiness Test: 1962 Edition	K-Gr.1	X				X		20 mins.	X		X	X		X						X	X				9	187
124. Lowrey-Lucier Reasoning Test Combination	Gr.5+	X	X	A,B		X		15-20 mins.			X	X		X							X				21	188
125. Marriage Role Expectation Inventory	Gr.9+	X	X	**			X	25-50 mins.	X		X	X		X	X								X		20	-
126. Meier Art Test	Gr.7-12	X	X			X	X	Not Given		X		X		X	X		X								7	188
127. Aesthetic Perception	Gr.9-16	X	X			X	X	Not Given		X		X		X	X		X								7	188
128. Art Judgment	Gr.7-12	X	X			X	X	Not Given		X		X		X	X		X								7	188
129. Metropolitan Achievement Tests: 1970 Edition	K.7-Gr.9.5	X		F,H	X	X	X	70-270 mins.	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	45	189
Primer	K.7-Gr.1.4	X		F,G,H	X	X	X	70 mins.	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	45	189
Primary I	Gr.1.5-2.4	X		F,G,H	X	X	X	90 mins.	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	45	189
Primary II	Gr.2.5-3.4	X		F,G,H	X	X	X	136 mins.	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	45	189
Elementary	Gr.3.5-4.9	X		F,G,H	X	X	X	180 mins.	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	45	189
Intermediate	Gr.5-6.9	X		F,G,H	X	X	X	270 mins.	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	45	189

** = Male, Female

Measures	Population	Type of Administration		Alternate Forms Available	Scoring		Availability of Scoring Services	Administration Time	Materials Available							Reliability Information Available	Validity Information Available	Scores Available								Publisher	Page of Test Description (if available)
		Group	Individual		Machine	Hand			Specimen Set	Reusable Test Booklet	One-time Test Booklet	Administrator's Manual	Interpretive Material	Practice Material	Standard Scores			Percentiles	Grade Equivalents	Age Equivalents	Stanines	Profiles/Scales	Raw Scores	Subscores	Not Indicated		
Advanced	Gr. 7-9.5	X		F, G, H	X	X	X	265 mins.	X	X	X	X	X	X	X	X	X	X	X		X	X	X	X		45	190
130. Metropolitan Readiness Tests: 1976 Edition	K-Gr. 1	X		P, Q	X	X		80-100 mins.	X	X	X	X	X	X	X	X		X		X	X	X	X		45	190	
Level I	K	X		P, Q	X	X		80-90 mins.	X	X	X	X	X	X	X	X		X		X	X	X	X		45	190	
Level II	Gr. 1	X		P, Q	X	X		90-100 mins.	X	X	X	X	X	X	X	X		X		X	X	X	X		45	190	
131. Minnesota Counseling Inventory	Gr. 9-12	X			X	X		50 mins.	X	X		X			X	X	X				X	X			45	190	
132. Minnesota Vocational Interest Inventory	Gr. 9+	X			X	X		45 mins.	X	X	X	X	X		X	X	X				X	X			45	191	
133. Mooney Problem Checklist: 1950 Edition	Gr. 7-12	X	X				X	35-50 mins.	X		X	X			X	X					X	X			45	191	
Junior High School Level	Gr. 7-9	X	X			X	X	35-50 mins.	X		X	X			X	X					X	X			45	-	
High School Level	Gr. 9-12	X	X			X	X	35-50 mins.	X		X	X			X	X					X	X			45	-	
134. Murphy-Durrell Reading Readiness Analysis	K. 8-Gr. 1.5	X				X		60 mins.	X		X	X			X	X		X		X	X	X			45	192	
135. Music Achievement Test	Gr. 4-12	X			X	X		30-35 mins.	X		X	X			X	X	X	X				X			24	192	
Level One	Gr. 4-12	X			X	X		35 mins.	X		X	X			X	X	X	X				X			24	192	

Measures	Population	Type of Administration		Alternate Forms Available	Scoring		Availability of Scoring Services	Administration Time	Materials Available							Scores Available										Publisher	Page of Test Description (if available)
		Group	Individual		Machine	Hand			Specimen Set	Reusable Test Booklet	One-time Test Booklet	Administrator's Manual	Interpretive Material	Practice Material	Reliability Information Available	Validity Information Available	Standard Scores	Percentiles	Grade Equivalents	Age Equivalents	Stanines	Profiles/Scales	Raw Scores	Subscores	Not Indicated		
Level Two	Gr. 4-12	X			X	X		35 mins.	X		X	X			X	X	X	X					X			24	192
Level Three	Gr. 4-12	X			X	X		30 mins.	X		X	X			X	X	X	X					X			24	193
Level Four	Gr. 4-12	X			X	X		35 mins.	X		X	X			X	X	X	X					X			24	193
136. Musical Aptitude Profile	Gr. 4-12	X			X	X		110 mins.	X	X		X	X		X	X							X	X		27	193
137. My Self Checklist	Gr. 4-9	X	X			X		Un-timed	X		X	X										X				22	193
138. NM Concepts of Ecology Test	Gr. 6-12	X				X			X		X	X			X			X			X		X	X		37	194
139. NM Consumer Mathematics Test	Gr. 9-12	X			X	X		20 mins.	X		X	X			X	X		X			X		X			37	194
140. NM Consumer Rights and Responsibilities Test	Gr. 9-12	X			X	X		20 mins.	X		X	X			X	X		X			X		X			48	-
141. National Achievement Tests	K-Gr. 14	X				X		40-45 mins.	X		X	X			X	X							X	X		48	195
142. American History-Government-Problems of Democracy	Gr. 8-14	X		A, B		X		40 mins.	X		X	X			X	X							X	X		48	195
143. Health Education: Knowledge and Application	Gr. 7-13	X		A, B		X		45 mins.	X		X	X			X	X		X					X	X		48	195
144. Health	Gr. 3-8	X		A, B		X		40 mins.	X		X	X			X				X				X	X		48	-
145. History and Civics	Gr. 3-6	X		A, B		X		40 mins.	X		X	X			X								X	X		48	-

Measures

Measures	Population	Type of Administration		Alternate Forms Available	Scoring		Availability of Scoring Services	Administration Time	Materials Available						Reliability Available	Validity Information Available	Scores Available							Publisher	Page of Test Description (if available)		
		Group	Individual		Machine	Hand			Specimen Set	Reusable Test Booklet	One-time Test Booklet	Administrator's Manual	Interpretive Material	Practice Material			Standard Scores	Percentiles	Grade Equivalents	Age Equivalents	Stanines	Profiles/Scales	Raw Scores			Subscores	Not Indicated
146. Nelson Biology Test: Revised Edition	Gr. 9-13	X			X	X	X	40 mins.	X	X	X	X			X	X	X	X		X		X	X			45	196
147. Nelson-Denny Reading Test: Forms C and D	Gr. 9+	X		C,D	X	X	X	30 mins.	X	X	X	X	X		X	X		X	X			X	X			27	196
148. Nelson Reading Test: Revised Edition	Gr. 3-9	X		A,B	X	X		30 mins.	X	X	X	X	X		X	X		X	X				X			27	196
149. New Mexico Career Education Test Series	Gr. 9-12	X		A,B				15-20 mins.	X		X	X			X	X		X			X		X			37	196
150. NM Attitude Toward Work Test	Gr. 9-12	X		A,B		X		15 mins.	X		X	X			X	X		X			X		X			37	197
151. NM Career Development Test	Gr. 9-12	X		A,B		X		20 mins.	X		X	X			X	X		X			X		X			37	197
152. NM Career Oriented Activities Checklist	Gr. 9-12	X		A,B		X		20 mins.	X		X	X			X	X		X			X		X			37	197
153. NM Career Planning Test	Gr. 9-12	X		A,B		X		20 mins.	X		X	X			X	X		X			X		X			37	197
154. NM Job Application Procedures Test	Gr. 9-12	X		A,B		X		20 mins.	X		X	X			X	X		X			X		X			37	197
155. NM Knowledge of Occupations Test	Gr. 9-12	X		A,B		X		15 mins.	X		X	X			X	X		X			X		X			37	197
156. Ohio Vocational Interest Survey	Gr. 8-12	X			X	X		60-90 mins.	X	X		X	X		X	X		X			X	X	X	X		45	198
157. Opinions, Attitudes, and Interests Survey	Gr. 11-16	X			X		X	40-60 mins.		X		X	X		X	X							X				
158. Oral School Attitude Test	K-Gr. 3	X	X			X		Un-timed	X		X	X			X	X							X	X		33	198

Measures	Population	Type of Administration		Alternate Forms Available	Scoring		Availability of Scoring Services	Administration Time	Materials Available						Reliability Information Available		Validity Information Available		Scores Available								Publisher	Page of Test Description (if available)
		Group	Individual		Machine	Hand			Specimen Set	Reusable Test Booklet	One-time Test Booklet	Administrator's Manual	Interpretive Material	Practice Material					Standard Scores	Percentiles	Grade Equivalents	Age Equivalents	Stanines	Profiles/Scales	Raw Scores	Subscores	Not Indicated	
159. Orleans-Hanna Algebra Prognosis Test	Gr. 7-11	X			X	X	X	40 mins.	X	X		X			X	X			X			X		X	X		45	198
160. Orleans-Hanna Geometry Prognosis Test	Gr. 8-11	X			X	X	X	40 mins.	X	X		X			X	X			X			X		X	X		45	199
161. Parent-Adolescent Communication Inventory	13 yrs.+	X	X			X		20 mins.	X		X	X	X		X	X			X					X			20	199
162. Personal Adjustment Inventory	Gr. 9+	X	X	**		X		Un-timed	X		X	X			X	X								X			6	199
163. Personal Orientation Inventory	Gr. 9+	X	X		X	X	X	20 mins.	X	X		X	X		X	X			X				X	X	X		16	200
164. Personal Values Abstract	Gr. 9+	X	X			X		15-20 mins.	X		X	X												X	X		12	200
165. Pictorial Study of Values	Gr. 9+	X				X		20 mins.	X		X	X			X			X						X	X		48	200
166. Piers-Harris Children's Self-Concept Scale	Gr. 3-12	X	X			X		15-20 mins.	X		X	X			X	X								X	X		13	201
167. Planning Career Goals	Gr. 8-12	X			X		X	278 mins.	X	X		X	X											X	X		9	201
168. Prescriptive Mathematics Inventory	Gr. 4-8	X	X		X		X	120-210 mins.	X	X		X	X	X									X	X	X		9	202
Orange Level	Gr. 4-5	X	X		X		X	120-150 mins.	X	X		X	X	X										X	X		9	202
Aqua Level	Gr. 5-6	X	X		X		X	120-150 mins.	X	X		X	X	X									X	X	X		9	202
Purple Level	Gr. 6-7	X	X		X		X	120-180 mins.	X	X		X	X	X									X	X	X		9	202

** = Boys, Girls

Measures	Population	Type of Administration		Alternate Forms Available	Scoring		Availability of Scoring Services	Administration Time	Materials Available						Reliability Information Available	Validity Information Available	Scores Available								Publisher	Page of Test Description (if available)
		Group	Individual		Machine	Hand			Specimen Set	Reusable Test Booklet	One-time Test Booklet	Administrator's Manual	Interpretive Material	Practice Material			Standard Scores	Percentiles	Grade Equivalents	Age Equivalents	Stanines	Profiles/Scales	Raw Scores	Subscores		
Level C	Gr. 7-8	X	X		X		X	180-210 mins.	X	X		X	X	X						X	X	X		9	202	
169. Prescriptive Reading Inventory	Gr. 1.5-6.5	X	X		X	X	X	165-195 mins.	X	X		X	X	X						X	X	X		9	203	
Level Red/A	Gr. 1.5-2.5	X	X		X	X	X	195 mins.	X	X		X	X	X						X	X	X		9	-	
Level Green/B	Gr. 2-3.5	X	X		X	X	X	180 mins.	X	X		X	X	X						X	X	X		9	-	
Level Blue/C	Gr. 3-4.5	X	X		X	X	X	180 mins.	X	X		X	X	X						X	X	X		9	-	
Level Orange/C	Gr. 4-6	X	X		X	X	X	165 mins.	X	X		X	X	X						X	X	X		9	-	
170. Primary Reading Profiles	Gr. 1-3	X			X	X	X	90 mins.	X	X	X	X	X		X	X		X	X	X	X	X		27	203	
Level I	Gr. 1-2	X			X	X	X	90 mins.	X	X	X	X	X		X	X		X	X	X	X		27	-		
Level II	Gr. 2-3	X			X	X	X	90 mins.	X	X	X	X	X		X	X		X	X	X	X		27	-		
171. Primary Self-Concept Inventory	Pre K-Gr. 6	X				X		Un-timed	X		X	X			X	X		X			X	X		33	203	
172. Principles of Democracy Test	Gr. 9-12	X		A, B	X	X		40 mins.	X	X		X			X	X		X			X	X		55	204	
173. Purdue High School English Test	Gr. 9-12	X		1, 2	X	X		36 mins.	X	X		X			X	X		X			X	X		27	-	
174. QUESTA	Gr. 9-12	X	X	I, II	X	X	X	40-60 mins.	X		X	X	X							X		X		1	204	

Measures	Population	Type of Administration		Alternate Forms Available	Scoring		Availability of Scoring Services	Administration Time	Materials Available						Reliability Information Available	Validity Information Available	Scores Available								Publisher	Page of Test Description (if available)	
		Group	Individual		Machine	Hand			Specimen Set	Reusable Test Booklet	One-time Test Booklet	Administrator's Manual	Interpretive Material	Practice Material			Standard Scores	Percentiles	Grade Equivalents	Age Equivalents	Stanines	Profiles/Scales	Raw Scores	Subscores			Not Indicated
175. Rating Scale of Vocational Values, Interests, and Aptitudes	Gr.9+	X	X			X		45 mins.	X	X		X	X				X				X	X	X		16	204	
176. Remote Associates Test: High School Form	Gr.9-12	X	X			X		40 mins.	X		X	X			X		X					X			27	205	
177. SRA Assessment Survey	Gr.1-12	X			X	X		150-270 mins.	X		X	X	X		X	X	X	X			X	X	X		55	206	
178. Iowa Test of Educational Development	Gr.9-12	X		X5, Y5	X	X	X	195 mins.	X	X		X	X		X	X	X	X			X	X	X	X		55	206
179. Multilevel Edition of the Achievement Series	Gr.4-9	X		E,F	X	X	X	270 mins.	X	X		X	X		X	X		X	X		X	X	X	X		55	206
180. Primary Edition of the Achievement Series	Gr.1-4	X		E,F	X	X	X	150 mins.	X	X		X	X		X	X		X	X		X	X	X	X		55	206
181. SRA Reading Record: Second Edition	Gr.6-12	X				X		45 mins.	X	X		X	X		X	X		X	X			X	X	X		55	207
182. SCORE: School Curriculum Objective Referenced Evaluation	K-Gr.9	X			X		X	Var-ies	X	X		X	X		X	X		X				X				64	205
183. Language Arts-Reading	K-Gr.9	X			X		X	Var-ies	X	X		X	X		X	X		X					X			64	205
184. Mathematics	K-Gr.9	X			X		X	Var-ies	X	X		X	X		X	X		X					X			64	205
185. Science	K-Gr.9	X			X		X	Var-ies	X	X		X	X		X	X		X					X			64	203
186. Social Science	K-Gr.9	X			X		X	Var-ies	X	X		X	X		X	X		X					X			64	206
187. STS Junior Inventory	Gr.4-8	X			X	X	X	35-40 mins.	X	X		X											X	X		54	207
188. STS Youth Inventory	Gr.7-12	X			X	X	X	45 mins.	X		X	X			X	X		X					X	X		54	207

Measures	Population	Type of Administration		Alternate Forms Available	Scoring		Availability of Scoring Services	Administration Time	Materials Available						Reliability Information Available	Validity Information Available	Scores Available								Publisher	Page of Test Description (if available)		
		Group	Individual		Machine	Hand			Specimen Set	Reusable Test Booklet	One-time Test Booklet	Administrator's Manual	Interpretive Material	Practice Material			Standard Scores	Percentiles	Grade Equivalents	Age Equivalents	Stanines	Profiles/Scales	Raw Scores	Subscores			Not Indicated	
189. San Diego County Inventory of Reading Attitude	Gr.1-4	X				X		Not Given	X		X	X			X	X					X		X				53	208
190. Sares-Sanders American Government Test	Gr.9-16	X		A,B	X	X		40 mins.	X	X		X			X	X									X		8	-
191. Scale of Socio-Egocentrism: Revised Edition, 1975	Gr.7-16	X				X		15-20 mins.				X	X										X	X			43	208
192. School Attitude Survey	Gr.4-6					X		Not Given	X		X	X			X	X					X	X	X				4	208
193. School Interest Inventory	Gr.7-12	X			X	X		20 mins.	X		X	X	X		X	X							X				27	208
194. School Inventory	Gr.9-12	X				X		15 mins.	X		X	X			X	X							X				12	209
195. School Motivation Analysis Test	12-18 yrs.	X	X	A,B	X	X		50-60 mins.	X	X	X	X												X	X		28	209
196. Seashore Measure of Musical Talent: 1956 Revision	Gr.4+	X			X	X		60 mins.	X	X		X			X	X		X						X	X		45	209
197. Self-Concept Adjective Checklist	K-Gr.8	X	X			X		Un-timed	X		X	X			X	X							X	X			47	210
198. Self-Concept and Motivation Inventory	Pre K-Gr. 12	X			X	X		Not Given	X		X	X			X						X		X	X			42	210
Preschool/Kindergarten	Pre K-K	X			X	X		Not Given	X		X	X			X						X		X	X			42	-
Early Elementary	Gr.1-3	X				X		Not Given	X		X	X			X						X		X	X			42	-
Later Elementary	Gr.3-6	X				X		Not Given	X		X	X			X						X		X	X			42	-

Measures	Population	Type of Administration		Alternate Forms Available	Scoring		Availability of Scoring Services	Administration Time	Materials Available							Scores Available							Publisher	Page of Test Description (if available)			
		Group	Individual		Machine	Hand			Specimen Set	Reusable Test Booklet	One-time Test Booklet	Administrator's Manual	Interpretive Material	Practice Material	Reliability Information Available	Validity Information Available	Standard Scores	Percentiles	Grade Equivalents	Age Equivalents	Stanines	Profiles/Scales			Raw Scores	Subscores	Not Indicated
Secondary	Gr. 7-12	X			X		Not Given	X		X	X			X					X		X	X		42	-		
199. Self-Directed Search	15 yrs.+	X	X		X		30-50 mins.	X		X	X	X		X	X						X	X		12	210		
200. Self-Directed Search: Form E	Gr. 4+	X	X		X		Not Given	X		X	X	X		X	X						X	X		12	211		
201. Self-Observation Scales	K-Gr. 6	X		A,B	X		20-25 mins.	X	X		X			X	X	X	X			X	X	X	X		39	211	
202. Self-Perception Inventory	Gr. 9-16	X	X		X		20-35 mins.	X		X	X			X	X	X	X			X	X	X	X		47	211	
203. Sequential Tests of Educational Progress: Original Series	Gr. 4-14	X			X	X	70-420 mins.	X	X		X	X		X	X	X					X			1	211		
Level 1	Gr. 13-14	X			X	X	70-420 mins.	X	X		X	X		X	X	X				X	X	X		1	-		
Level 2	Gr. 10-12	X			X	X	70-420 mins.	X	X		X	X		X	X	X				X	X	X		1	-		
Level 3	Gr. 7-9	X			X	X	70-420 mins.	X	X		X	X		X	X	X				X	X	X		1	-		
Level 4	Gr. 4-6	X			X	X	70-420 mins.	X	X		X	X		X	X	X				X	X	X		1	-		
204. Sequential Tests of Educational Progress: Series Two	Gr. 4-14	X		A,B	X	X	40-325 mins.	X	X		X	X		X	X	X				X	X	X		1	-		
Level 1	Gr. 13-14	X		A,B	X	X	40-325 mins.	X	X		X	X		X	X	X			X		X	X		1	212		
Level 2	Gr. 10-12	X		A,B	X	X	40-325 mins.	X	X		X	X		X	X	X			X		X	X		1	-		
								X	X		X	X		X	X	X			X		X	X		1	-		

80

Measures	Population	Type of Administration		Alternate Forms Available	Scoring			Administration Time	Materials Available						Reliability Information Available	Validity Information Available	Scores Available								Publisher	Page of Test Description (if available)
		Group	Individual		Machine	Hand	Availability of Scoring Services		Specimen Set	Reusable Test Booklet	One-time Test Booklet	Administrator's Manual	Interpretive Material	Practice Material			Standard Scores	Percentiles	Grade Equivalents	Age Equivalents	Stanines	Profiles/Scales	Raw Scores	Subscores		
Level 3	Gr. 7-9	X		A,B	X	X	X	40-290 mins.	X	X		X	X		X	X		X			X	X		1	-	
Level 4	Gr. 4-6	X		A,B	X	X	X	40-290 mins.	X	X		X	X		X	X		X			X	X		1	-	
205. Similes Test	Gr. 4+	X		I, II		X		15 mins.			X	X			X	X		X			X			49	214	
206. Situational Preference Inventory	Gr. 9+	X			X	X		10-15 mins.			X	X			X	X					X	X		19	214	
207. Smith-Sturgeon Conditional Reasoning Inventory	6-9 yrs.		X			X		40-60 mins.			X		X		X	X					X	X		14	215	
208. Social and Prevocational Information Battery	Gr.7-12	X			X	X		15-25 mins.	X	X	X	X			X	X		X			X	X		9	215	
209. Social Intelligence Test	Gr. 9+	X				X		45+ mins.	X		X	X			X	X					X	X		11	-	
210. Stanford Achievement Tests: 1973 Edition	Gr.1.5-9.5	X		A,B	X	X	X	170-320 mins.	X	X	X	X	X	X	X	X		X	X		X		X	X	45	215
Primary Level I	Gr.1.5-2.4	X		A,B	X	X	X	170 mins.	X	X	X	X	X	X	X	X		X	X		X		X	X	45	216
Primary Level II	Gr.2.5-3.4	X		A,B	X	X	X	260 mins.	X	X	X	X	X	X	X	X		X	X		X		X	X	45	216
Primary Level III	Gr.3.5-4.4	X		A,B	X	X	X	295 mins.	X	X	X	X	X	X	X	X		X	X		X		X	X	45	216
Intermediate Level I	Gr.4.5-5.4	X		A,B	X	X	X	320 mins.	X	X	X	X	X	X	X	X		X	X		X		X	X	45	216
Intermediate Level II	Gr.5.5-6.9	X		A,B	X	X	X	320 mins.	X	X	X	X	X	X	X	X		X	X		X		X	X	45	216

Measures	Population	Type of Administration		Scoring			Administration Time	Materials Available							Scores Available										Publisher	Page of Test Description (if available)
		Group	Individual	Alternate Forms Available	Machine	Hand		Specimen Set	Reusable Test Booklet	One-time Test Booklet	Administrator's Manual	Interpretive Material	Practice Material	Reliability Information Available	Validity Information Available	Standard Scores	Percentiles	Grade Equivalents	Age Equivalents	Stanines	Profiles/Scales	Raw Scores	Subscores	Not Indicated		
Advanced Level	Gr. 7-9.5	X		A, B	X	X	X 260 mins.	X	X	X	X	X	X	X	X		X	X		X		X	X		45	217
211. Stanford Achievement Tests: High School Battery	Gr. 9-12	X		W, X	X	X	X 320 mins.	X	X	X	X	X		X	X		X			X	X	X	X		45	-
212. Stanford Diagnostic Arithmetic Test: 1966 Edition	Gr. 2.5-8.5	X		W, X		X	170-210 mins.	X	X	X	X	X		X	X		X	X		X	X	X	X		45	217
Level I	Gr. 2.5-4.5	X		W, X		X	170 mins.	X	X	X	X	X		X	X		X	X		X	X	X	X		45	-
Level II	Gr. 4.5-8.5	X		W, X		X	210 mins.	X	X	X	X	X		X	X		X	X		X	X	X	X		45	-
213. Stanford Diagnostic Mathematics Test: 1976 Edition	Gr. 1.5-13	X		A, B	X	X	X 90-110 mins.	X	X	X	X	X		X	X		X	X		X	X	X	X		45	217
Red Level	Gr. 1.5-4.5	X		A, B	X	X	X 110 mins.	X	X	X	X	X		X	X		X	X		X	X	X	X		45	-
Green Level	Gr. 3.5-6.5	X		A, B	X	X	X 95 mins.	X	X	X	X	X		X	X		X	X		X	X	X	X		45	-
Brown Level	Gr. 5.5-8.5	X		A, B	X	X	X 100 mins.	X	X	X	X	X		X	X		X	X		X	X	X	X		45	-
Blue Level	Gr. 7.5-13	X		A, B	X	X	X 90 mins.	X	X	X	X	X		X	X		X	X		X	X	X	X		45	-
214. Stanford Diagnostic Reading Test: 1976 Edition	Gr. 1.5-9.5	X		A, B	X	X	X 103-130 mins.	X	X	X	X	X	X	X	X		X	X		X	X	X	X		45	217
Red Level	Gr. 1.5-3.5	X		A, B	X	X	X 130 mins.	X	X	X	X	X	X	X	X		X	X		X	X	X	X		45	-
Green Level	Gr. 2.5-5.5	X		A, B	X	X	X 145 mins.	X	X	X	X	X	X	X	X		X	X		X	X	X	X		45	-

Measures	Population	Type of Administration		Alternate Forms Available	Scoring		Availability of Scoring Services	Administration Time	Materials Available						Reliability Information Available	Validity Information Available	Scores Available								Publisher	Page of Test Description (if available)
		Group	Individual		Machine	Hand			Specimen Set	Reusable Test Booklet	One-time Test Booklet	Administrator's Manual	Interpretive Material	Practice Material			Standard Scores	Percentiles	Grade Equivalents	Age Equivalents	Stanines	Profiles/Scales	Raw Scores	Subscores		
Brown Level	Gr.4.5-9.5	X		A,B	X	X	X	103 mins.	X	X	X	X	X	X	X		X	X		X	X	X	X		45	-
215. Stanford Early School Achievement Test	K.1-Gr. 1.8	X			X	X		110-140 mins.	X	X	X	X	X	X	X		X			X	X	X	X		45	218
Level I	K.1-Gr. 1.1	X			X	X		110 mins.	X	X	X	X	X	X	X		X			X	X	X	X		45	-
Level II	Gr.1.1-1.8	X			X	X		140 mins.	X	X	X	X	X	X	X		X			X	X	X	X		45	-
216. Stanford Tests of Academic Skills (TASK)	Gr.8-13	X		A,B	X	X		80 mins.	X	X	X	X			X	X		X		X		X	X		45	218
Level I	Gr.8-10	X		A,B	X	X	X	80 mins.	X	X	X	X			X	X		X		X		X	X		45	-
Level II	Gr. 11-13	X		A,B	X	X	X	80 mins.	X	X	X	X			X	X		X		X		X	X		45	-
217. Strong-Campbell Interest Inventory	16 yrs.+	X	X		X		X	25-35 mins.	X	X		X	X		X	X	X					X	X		57	219
218. Student Evaluation Scale	Gr.1-12	X				X		5 mins.	X		X	X											X		47	-
219. Subsumed Abilities Test	Gr. 6+	X				X		30 mins.	X		X	X			X	X		X				X			35	-
220. Survey of Interpersonal Values	Gr. 9+	X	X			X		15 mins.	X		X	X			X	X		X				X	X		55	219
221. Survey of Personal Values	Gr. 0+	X	X			X		15 mins.	X		X	X			X	X		X				X	X		55	219
222. Survey of School Attitudes	Gr. 1-8	X	X	A,B	X	X		20-30 mins.	X	X		X			X	X		X				X	X		45	220

Measures	Population	Type of Administration		Alternate Forms Available	Scoring		Availability of Scoring Services	Administration Time	Materials Available						Reliability Information Available	Validity Information Available	Scores Available								Publisher	Page of Test Description (if available)	
		Group	Individual		Machine	Hand			Specimen Set	Reusable Test Booklet	One-time Test Booklet	Administrator's Manual	Interpretive Material	Practice Material			Standard Scores	Percentiles	Grade Equivalents	Age Equivalents	Stanines	Profiles/Scales	Raw Scores	Subscores			Not Indicated
223. Survey of Study Habits and Attitudes: High School Form	Gr.7-12	X		C,H	X	X	X	20-25 mins.	X	X		X			X	X		X					X	X		45	220
224. Taylor-Helmstadter Pair Comparison Scale of Aesthetic Judgment	4 yrs.+	X				X		Not Given		X		X			X	X							X			61	220
225. Taylor-Johnson Temperament Analysis: Secondary Edition	Gr.7-12	X				X		Un-timed	X		X	X	X		X	X					X	X	X			46	-
226. Tennessee Self-Concept Scale	17 yrs.+	X	X		X	X		Not Given	X	X		X	X		X	X							X	X		13	221
227. Test of Creative Potential	7 yrs.+	X		A,B		X		20-30 mins.	X		X	X			X	X		X		X			X	X		37	221
228. Test of Elementary Economics	Gr. 4-6	X			X	X		Un-timed	X		X	X					X						X	X		30	222
229. Test of Social Insight: Youth Edition	Gr.6-12	X	X			X		Not Given	X		X	X			X	X								X		35	222
230. Test of Understanding in Personal Economics	Gr.9-12	X			X	X		45 mins.	X		X	X			X			X					X			30	222
231. Tests of Academic Progress: Form S	Gr.9-12	X			X	X	X	270 mins.	X	X	X	X	X		X	X	X	X			X		X	X		27	-
232. Tests of Basic Experiences	Pre K-Gr.1	X			X	X	X	25 mins.	X	X		X	X		X	X	X	X		X			X	X		9	223
Level K	Pre K-K	X			X	X	X	25 mins.	X	X		X	X		X	X	X	X		X			X	X		9	-
Level L	Gr.1	X			X	X	X	25 mins.	X	X		X	X		X	X	X	X		X			X	X		9	-
233. Thinking Creatively with Sounds and Words	Gr.3-12	X	X	IA, IB, IIA, IIB		X		30-35 mins.	X		X	X	X		X	X	X						X			41	223

Measures	Population	Type of Administration		Alternate Forms Available	Scoring		Availability of Scoring Services	Administration Time	Materials Available						Reliability Available	Validity Information Available	Scores Available								Publisher	Page of Test Description (if available)	
		Group	Individual		Machine	Hand			Specimen Set	Reusable Test Booklet	One-time Test Booklet	Administrator's Manual	Interpretive Material	Practice Material			Standard Scores	Percentiles	Grade Equivalents	Age Equivalents	Stanines	Profiles/Scales	Raw Scores	Subscores			Not Indicated
234. Thomas Self-Concept Values Test	3-9 yrs.	X				X		Un-timed	X		X	X			X	X	X					X	X	X		62	223
235. Thurstone Interest Schedule	Gr. 9+	X				X		10 mins.	X	X	X	X			X	X							X	X		45	224
236. Torrance Tests of Creative Thinking	K+	X	X	A,B		X		65 mins.	X		X	X			X	X								X		41	224
237. Watson-Glaser Critical Thinking Appraisal	Gr. 9+	X		Ym, Zm	X	X	X	50 mins.	X	X		X			X	X		X			X		X	X		45	225
238. Wide Range Interest and Opinion Test	K+	X	X		X	X		50-60 mins.	X	X		X			X	X							X	X		26	225
239. Woodcock Reading Mastery Test	K-Gr.12		X			X		20-30 mins.	X	X		X	X	X	X	X	X	X	X	X			X	X		3	225
240. Work Values Inventory	Gr. 7+	X			X		X	15 mins.	X	X		X			X	X							X	X		27	226
241. Written School Attitudes Test	Gr. 4-6	X	X	Eng. Spa.		X		30 mins.	X		X	X			X	X							X	X		33	226
242. Youth Research Survey	Gr.9-13	X				X		Not Given	X		X	X			X	X	X	X					X	X		66	227
243. Zimmerman-Sanders Social Studies Test	Gr. 7-8	X		A,B			X	30 mins.	X		X	X			X	X			X				X			8	-

APPENDIX A-3 - Compendium of Descriptions of Tests (External Measures)

The descriptions that follow pertain to tests that are relevant to one or more of the New Jersey State Educational Outcome Goals. A test is designated as appropriate to a specific goal if the publisher has indicated that it is specifically designed to assess attainment of that goal (or similar construct), or if it includes subtests or items that could be used for that purpose. Although this compendium does not include all the tests listed in the Matrix of Test Descriptions, full representation across all outcome goals has been maintained.

The tests were identified through a search of the holdings of the ETS Test Collection. An effort was made to concentrate on commercially available instruments. However, in instances where an unpublished instrument was relevant and accessible, it was included.

The first part of the description consists of the basic identifying information: test title, author(s), publication or copyright date, the population for which the test is appropriate, and the test publisher or other source from which the test can be obtained. This is followed by an abstract consisting of a brief statement of the instrument's objective, a listing of available subtests or scores, an indication of any unusual characteristics (e.g., special administrator qualifications or instructions in Spanish), and a description of the technical information available in the test manual or technical report.

These descriptions are intended to be informative, not evaluative. Because an effort was made to cover a wide variety of tests, many instruments have not been included in this listing. This, too, should not be regarded as an evaluative statement. Inclusion of a test in this listing is not intended as a recommendation for its use. Finally, it is recommended that each individual judge the value of a given test in terms of his own needs and objectives. This can only be accomplished through a careful review of the actual test and supporting materials.

AAHPER Cooperative Health Education Tests; c1971; Grades 4-9; Addison-Wesley Publishing Company, Inc.

Designed to measure knowledge and understanding of concepts taught in health education classes, the test covers such areas as: consumer health, community health, international health, disease and disorders, personal health, sex education (for grades 7-9 only), growth and development, nutrition, mental health, drug use and abuse, and safety and first aid. Two levels of the test are available.

ACT Career Planning Program (CPP); c1976; Ages 16 and Over; American College Testing Program.

This is a guidance-oriented system designed to help students identify and explore personally occupations and educational programs. The student report and booklet help students relate information about their interests, experiences, and abilities to the world of work. The CPP collects four major types of information contained within four sections of the instrument. These sections include a Vocational Interest Profile (Business Contact, Business Detail, Trades, Technology, Science, Health, Creative Arts, Social Services); Experience Scales summarizing the work-related experiences reported by the student in the same basic areas measured by the interest scales; Ability Measures (Mechanical Reasoning, Numerical Skills, Space Relations, Reading Skills, Language Usage, Clerical Skills); and a Student Information Section where students describe a number of factors about themselves relevant to vocational and educational planning.

Adjustment Inventory: Revised Student Form by Hugh M. Bell; c1962; Grades 9-16; Consulting Psychologists Press, Inc.

This device elicits information concerning how the individual thinks and feels about his family relationships; his functioning body, and his friends and acquaintances, and how well he is able to fulfill the roles that society expects of him. It yields scores for six areas of personal and social adjustment: Home Adjustment, Health Adjustment, Submissiveness, Emotionality, Hostility, and Masculinity-Femininity. The inventory is self-administering and untimed. It can be hand scored or machine scored (IBM

805). Percentile norms were drawn from seven high schools and five colleges throughout the country. However, the author recommends that institutions develop their own normative data. Reliability coefficients were determined by correlating the odd-even items and applying the Spearman-Brown formula. Corrected coefficients for the six scales range from .80 to .89. Cross-validation studies of the six scales and construct validity investigation of four scales are discussed in the manual.

American Political Behavior Achievement Test by John J. Patrick and Allen D. Glenn; c1974; Grades 9-12; Ginn and Company.

This two-part test is designed for use with civics and government courses that are intended to teach knowledge of political affairs in the U.S. and particular intellectual skills associated with higher-level learning about political life. Part I, Political Knowledge, covers various facets of American government and politics: elections and voter behavior; roles of governmental leaders; roles of unofficial political specialists; socio-cultural contexts of political behavior; and governmental structure and organization. Part II, Political Science Skills, measures various intellectual skills: reading contingency tables; reading graphs; making inductive inferences; making deductive inferences; distinguishing factual statements, value statements, and definitional statements; evaluating questionnaire items; evaluating sampling procedures; evaluating definitions; and conceptualizing. Three reliability studies were conducted using the Spearman-Brown test. Reliability coefficients for Part I were .88, .92, and .93; for Part II .77, .88, and .90. Validity is not discussed.

Analysis of Readiness Skills, Reading and Mathematics by Mary C. Rodrigues, William H. Vogler, and James F. Wilson; c1969-72; Grades K-1; Houghton Mifflin Company.

This short test for children beginning formal schooling provides schools with a means of determining the most appropriate time for each child to enter reading and mathematics programs. The analysis consists of three brief, uncomplicated tests: Visual Perception of Letters, Letter Identification, and Mathematics. The tests may also be administered in Spanish with the results for Spanish-speaking children indicating their readiness for placement in an ESL (English as a Second Language) program. Odd-even, split-half reliability for the total test corrected with the Spearman-Brown formula yielded a coefficient of .928. The Kuder-Richardson 20 reliability coefficient computed was .912.

Animal Crackers: A Test of Motivation to Achieve by Dorothy C. Adkins and Bonnie L. Ballif; c1973; Preschool-Grade 1; CTB/McGraw-Hill.

This test focuses on five aspects of achievement-oriented behavior: school enjoyment, self-confidence, purposiveness, instrumental activity, and self-evaluation. It employs an objective-projective technique that requires the child to choose between alternative behaviors or attitudes, described verbally, that show differences in motivation. The test may be individually or group-administered. For comparison with a reference group sample for kindergarten or first grade, raw scores may be converted to percentile ranks. However, the author recommends the development of local norms. The K-R 20 reliability coefficients for the five components range from .65 to .82; the K-R 20 reliability coefficient is .90 for the total test. The author claims evidence for content validity, construct validity, and criterion-related validity. Further validity studies are in progress.

Arlin-Hills Attitude Survey by Marshall N. Arlin, David Hills, N. Arlin; c1974; Kindergarten-Grade 12; Marshall N. Arlin.

This device is designed to elicit information on students' school-related attitudes. The survey consists of four instruments that deal with attitudes toward teachers, learning processes, language arts, and mathematics. Each of the tests is available at three levels: primary (k-3), elementary (4-6), and high school (7-12). The survey is designed to be administered and interpreted on a group basis and is intended primarily for diagnostic, criterion-referenced use and not for comparative, normative use. However, normative information is provided for comparative purposes and is based on a selected sample of 14,000 North Carolina students from 50 schools throughout the state.

Assessment of Career Development (ACD); c1972-74; Grades 8-11; Houghton Mifflin Company.

A new kind of guidance assessment instrument designed for use with eighth through eleventh grade students as part of a school's developmental career guidance program, the ACD focuses on core aspects of career development that can be objectively measured through use of standardized group assessment procedures. Scales in the ACD are structured around the

following components of career development: Occupational Awareness which includes Occupational Knowledge and Exploratory Occupational Experiences; Self-Awareness including Preferred Job Characteristics, Career Plans, and Perceived Needs for Help; and Career Planning and Decision Making concerned with Career Planning Knowledge and Career Planning Involvement. The Spearman-Brown formula used to obtain reliability estimates yielded coefficients ranging from .71 to .93. Content validity is discussed in the handbook.

Attitude Questionnaire by Gary A. Davis, Susan E. Houtman, Thomas E. Warren, and William E. Roweton; c1969; Grades 7-8; Wisconsin Research and Development Center for Cognitive Learning.

This instrument is designed to detect individuals' feelings about creativity in general and in relation to themselves. More specifically, the questionnaire seeks to determine individuals' appreciation for and receptivity to imaginative ideas, their awareness of creativity and innovation, and their perception of their own creative capacity. Students are advised to be honest as there are no "right" or "wrong" answers.

Attitude Toward Classroom Atmosphere (ATCA); 1974; Grades 1-6; Educational Improvement Center.

This self-report instrument designed to measure children's attitudes toward their classroom environment and toward their teachers is available in two formats: primary edition, grades 1-3, and elementary edition, grades 4-6. The attitude statements may or may not be read aloud to the students, depending on the age/range level and on the reading ability of the students. Test-retest reliability coefficients ranged from .19-.78 for grades 2 and 5 over a four-week interval. Content validity was examined in the pilot stage and concurrent validity was measured by correlating total scores on the ATCA with other measures given to the same children. Because attitude scales for children tend to be low in reliability, studies of the validity of the ATCA are still in the initial tentative stages.

Barclay Classroom Climate Inventory (BCCI) by James R. Barclay, Lisa K. Barclay, Calvin Catterall, David A. Santaro, William C. Stillwell III, and George Tapp; c1967-72; Grades 2-6; Educational Skills Development, Inc.

A multiple-measurement device for assessing individual differences and the characteristics of the classroom climate, the BCCI measures social interaction and expectation variables for elementary school children. There are 42 short scales that measure self-competency skills, peer judgments, vocational awareness, behavioral reinforcers, and teacher expectations. The inventory provides relevant diagnostic information for school personnel. Test-retest reliability coefficients range from .39 to .80 over the various scales in the inventory. Internal consistency coefficients range from .50 to .74. Cross-validation studies with other tests have been performed in the areas of personality, vocational assessment, intelligence, and achievement.

Basic Skills in Arithmetic by William L. Wrinkle, Juanita Sanders, and Elizabeth H. Kendel; c1945; Grades 6-12; Science Research Associates, Inc.

Designed to measure command of 43 fundamental skills in arithmetical operations involving whole numbers, fractions, decimals, and percents, the test reveals to the student and the teacher proficiency or lack of competence in the use of groups of skills as well as the use of specific skills. The test results provide a basis for the determination of group and individual instructional needs, and for the selection and counseling of students enrolling in courses and activities requiring basic skills in arithmetic. Reliability coefficients computed by the use of the Kuder-Richardson formula ranged from .86 to .98 for the sixth through twelfth grade.

Boehm Test of Basic Concepts (BTBC), 1969 Edition by Ann E. Boehm; c1967-71; Grades K-2; The Psychological Corporation.

This test is designed to measure children's mastery of concepts considered necessary for achievement in the first years of school. Two forms of the test are available along with four concept scales: Space, Quantity, Time, and Miscellaneous. The BTBC provides a means for diagnosing the attainment

of these specific conceptual terms for instructional purposes. After the strengths and weaknesses of each pupil are ascertained, the teacher may help the pupil to acquire the concepts necessary to perform tasks required to use the instructional materials. Split-half reliability coefficients, based on the correlation of scores on odd-numbered items with scores on even-numbered items, corrected by the Spearman-Brown formula for Form A were .90, .85, and .81 for the kindergarten, grade 1, and grade 2 samples respectively. For Form B, the coefficients were .83, and .87 for the same grades respectively. Content validity of the BTBC has been evaluated.

Brown-Carlson Listening Comprehension Test by James I. Brown, and G. Robert Carlson; c1953-55; Grades 9-13 and Adults; The Psychological Corporation.

The test has been constructed to measure the ability of students to comprehend spoken language. Two comparable forms of the test are available (Am and Bm), each comprising 76 test items. These items are grouped into five parts, each measuring an important listening skill. Subtests include: Immediate Recall, Following Directions, Recognizing Transitions, Recognizing Word Meanings, and Lecture Comprehension. Split-half reliability coefficients for Form Am were .86 and .84-.90 for grades 10 and 12 respectively. Reliability obtained by correlating scores on alternate forms of the test yielded a median estimate of .78. Validity data were not available from the manual.

Bryant-Schwan Design Test by Anusa S. Bryant and LeRoy B. Schwan; c1971-73; Preschool-Grade 12; Campus Publishers.

The test is geared toward measuring achievement in design knowledge of mentally retarded children, although it can be used with normal children for the same purpose. It can also be used as a diagnostic tool to identify what elements of art are known or not known by children. The five elements of design measured include line, shape, value, color, and texture. Twelve items for each of the five elements make up each of the two parts of the test. The first part consists of matching items; the second part consists of identification items. Test-retest reliability studies were conducted with a range of coefficients presented in the manual. Validity of the items were analyzed in relation to several variables.

California Achievement Tests (CAT): 1970 Edition by Ernest W. Tiegs and Willis W. Clark; c1957-70; Grades 1.5-12; CTB/McGraw-Hill.

This instrument is designed for the measurement, evaluation, and analysis of school achievement. The emphasis is on content and objectives in the basic curricular areas of reading, mathematics, and language. The intended measurement is one of performance in these curricular areas. The battery is comprised of five levels for the various grades as follows: Level I, grades 1.5-2; Level II, grades 2-4; Level III, grades 4-6; Level IV, grades 6-9; Level V, grades 9-12. Subtests of the basic skill areas assessed for the five levels include: Reading (Vocabulary and Comprehension); Mathematics (Computation and Concepts, and Problems); and Language (Mechanics: Capitalization and Punctuation, Usage and Structure, and Spelling). K-R 20 reliability coefficients range from .73 to .96 for Level I; .83 to .96 for Level II; .63 to .95 for Level III; .68 to .95 for Level IV; and .66 to .96 for Level V. The technical report containing extensive information on the CAT, as well as reports of a variety of validity studies, is scheduled to appear in the future.

California Life Goals Evaluation Schedules (CLGES) by Milton E. Hahn; c1966-69; Ages 16 and Over; Western Psychological Services.

This inventory is designed to differentiate between the motivations behind interests and life goals. The schedules relate attitudes in the areas of economic, social, and political thinking to the long-range planning of activities in significant life areas such as career selections, retirement planning, identification of avocations, relating to specific groups or individuals, and the recognition of acceptable life values and philosophical or religious positions. The 10 classifications and definitions identified as schedules in the CLGES include: Esteem, Profit, Fame, Power, Leadership, Security, Social Service, Interesting Experiences, Self-Expression, and Independence. Kuder-Richardson and Spearman-Brown (odd-even) coefficients of reliability for Forms A-B have been reported ranging from .87 to .98. For Form C the range is from .82-.89. Test-retest studies for Form D-M produced reliabilities ranging from .71 to .86. Content, concurrent, predictive, construct, and face validity information is available.

California Occupational Preference Survey (COPS) by Robert R. Knapp, Bruce Grant, and George D. Demos; c1966-71; Grades 9-16 and Adults; Educational and Industrial Testing Service.

This interest inventory provides systematic information regarding students' interests in groups of meaningfully related occupations. The COPS is intended as an aid to young men and women in making their occupational choices. Scores may be obtained for 14 occupational scales representing clusters of occupations at two levels within six major groupings plus scales for two additional groupings. The scales include: Science: Professional; Science: Skilled; Technical: Professional; Technical: Skilled; Outdoor; Business: Professional; Business: Skilled; Clerical; Linguistic: Professional; Linguistic: Skilled; Aesthetic: Professional; Aesthetic: Skilled; Service: Professional; and Service: Skilled. Split-half reliability coefficients range from .86 to .95. Validity data are available.

Career Awareness Inventory by La Verna M. Fadale; c1974; Grades 4-12; Scholastic Testing Service.

This instrument is a tool for classroom teachers and other educators interested in assessing career awareness as demonstrated by their students. The element of career awareness includes the knowledge, social attitudes, personal experiences, and contemplations elementary school children demonstrate about careers and occupations. The seven subparts of the inventory include Identity, Training, Models, Function, Prestige, Clusters, and Characteristics. Split-half, odd-even reliability corrected by the Spearman-Brown formula yielded a coefficient of .80. Validity data is available in the teacher's manual.

Career Development Inventory (CDI): Form I by Donald E. Super, David J. Forrest, Martin J. Bohn, Jr., Jean Pierre Jordan, Richard H. Lindeman and Albert S. Thompson; c1971-72; Grades 7-16; Donald E. Super.

An objective inventory measuring vocational maturity, the CDI assesses three of the important desired outcomes of career development and vocational guidance programs: the development of a planning orientation

toward a career, familiarity with and use of the resources that can be useful in vocational exploration, and knowledge of occupations and of career decision-making principles. Subscales include: Planning Orientation, Resources for Exploration, and Information and Decision Making. Test-retest reliability coefficients range from .71 to .85 for the three subscales. Content, criterion-related, and construct validity have been examined.

Career Maturity Inventory (CMI) by John O. Crites; c1973-76; Grades 6-12; CTB/McGraw-Hill.

This inventory is designed to measure the maturity of attitudes and competencies that are critical in realistic career decision making. The CMI provides two types of measures: the Attitude Scale and the Competence Test. The Attitude Scale elicits the feelings, the subjective reactions, and the dispositions that the individual has toward making a career choice and entering the world of work. The Competence Test measures the more cognitive variables involved in choosing an occupation. Subscales include: Knowing Yourself, Knowing About Jobs, Choosing a Job, Looking Ahead, and What Should They Do? Kuder-Richardson Formula 20 reliability coefficients range from .70 to .84 for grades 6-12. Content, criterion-related, and construct validity data are available.

Career Planning Program, Grades 8-11; c1974; Grades 8-11; Houghton Mifflin Company.

Developed to help schools meet the career guidance needs of their students, the program is designed to assess student experiences, interests, and abilities, and to aid the student in identifying and exploring career options. CPP 8-11 includes two key components. Booklet 1, Exploring: You and Your Career, introduces basic career-development concepts and career-planning resources and contains the experience and interest inventories. Booklet 2, Ability Measures, contains tests to assess students' aptitudes in six skill areas: Mechanical Reasoning, Numerical Skills, Space Relations, Reading Skills, Language Usage, and Clerical Skills. The program may also be adapted

for use as a mini-course in career guidance. Coefficient alpha was used as an internal estimate of reliability for the interest and experience scales while the Kuder-Richardson Formula 20 was used for the ability measures. The reliability coefficients for the various scales range from .71 to .92. Criterion-related validity and construct validity are discussed in the handbook.

The Child Behavior Rating Scale (CBRS) by Russel N. Cassel; c1962; Ages 3-9; Western Psychological Services.

This scale is a psychological instrument developed and standardized for the objective assessment of personality adjustments of preschool and primary pupils. The 78 CBRS items, each descriptive of some aspect of child behavior, are classified into five adjustment areas. These areas include: Self Adjustment, Home Adjustment, Social Adjustment, School Adjustment, and Physical Adjustment. Using the Spearman-Brown formula on odd-even CBRS items a reliability coefficient of .87 was computed for typical children. A coefficient of .60 was computed for maladjusted children. Ratings made on two separate occasions by parents and teachers produced coefficients of .91 and .74 respectively. Construct, face, and status validity data are available.

CIRCUS; c1974; Preschool-Kindergarten; Addison-Wesley Publishing Company, Inc.

CIRCUS was developed to provide a comprehensive selection of measures and services for use in diagnosing the instructional needs of individual children and in monitoring and evaluating early education programs. It consists of 17 instruments. The 14 direct child measures include: What Words Mean (receptive vocabulary), How Much and How Many (quantitative concepts), Look Alikes (visual discrimination), Copy What You See (perceptual-motor coordination), Finding Letters and Numbers (letter and numeral recognition), Noises (discrimination of real world sounds), How Words Sound (auditory discrimination), How Words Work (aspects of functional language), Listen to the Story (comprehension and interpretation of oral language), Say and Tell (productive language), Do You Know...? (general information), See and Remember (visual and associative memory), Think It Through (problem solving), and Make a Tree (divergent pictorial production). There are also three teacher-reporting instruments: Activities Inventory (children's observed interests), CIRCUS Behavior Inventory (children's responses to test situations), and Educational Environment Questionnaire (aspects of school and classroom environment). Individual measures can be selected for their relevance to specific goals of a program, or several measures can be used to provide comprehensive assessment.

Comprehensive Career Assessment Scale: Grades 3-12 by Steve L. Jackson and Peggy M. Goulding; c1974; Grades 3-12; Learning Concepts, Inc.

The test, designed to yield an index of the student's familiarity with and interest in each of the 15 U.S.O.E. occupational clusters, is available on two levels: Grades 3-7 and Grade 8-12. The dimensions of interest and familiarity are assessed through 15 career clusters which include the following: Agribusiness and Natural Resources, Business and Office, Communications and Media, Construction, Consumer and Homemaking Education, Environment, Fine Arts and Humanities, Health, Hospitality and Recreation, Manufacturing, Marine Science, Marketing and Distribution, Personal Service, Public Service, and Transportation. The alpha reliability coefficient computed for both interest and familiarity was .92 for grades 3-7. Grades 8-12 yielded alpha coefficients of .95 and .93 for familiarity and interest respectively.

Comprehensive Tests of Basic Skills: Forms S and T; c1973-75; Grades K-12; CTB/McGraw-Hill.

This instrument is designed to measure the level of attainment of language, number, and problem-solving skills required for academic study and for everyday needs outside of school. The items are classified by a taxonomy of intellectual skills across content areas of reading, language, mathematics, reference skills, science, and social studies. The CTBS contains seven levels spanning grades K-12 as follows: Level A, grades K.0-1.3; Level B, K.6-1.9; Level C, grades 1.6-2.9; Level 1, grades 2.5-4.9; Level 2, grades 4.5-6.9; Level 3, grades 6.5-8.9; Level 4, grades 8.5-12.9.

Level A is a preinstructional or readiness instrument for children in kindergarten or children in grade 1 who have not attended kindergarten. The total prereading score comprises seven test scores: Letter Forms, Letter Names, Listening for Information, Letter Sounds, Visual Discrimination, Language, and Sound Matching. There is also a mathematics test that assesses intuitive bases for number systems, operations, and measurements.

Level B is designed for students who have received approximately a year of instruction in either kindergarten or grade 1. Subtests include: Word Recognition, Letter Sounds, Reading Comprehension, Language, Mathematics Computation, and Mathematics Concepts and Applications.

Level C is given in the second grade and forms a measurement transition between tests for educationally naive students (Levels A and B at grades K and 1) and tests for students at grade 2.5 and above who can be evaluated appropriately by Level 1 through 4. The subtests and scores are similar to those in the upper levels, but format is more appropriate to very young students.

Levels C, 1, 2, 3, and 4 are composed of 10 separate subtests that combine to yield Total Reading, Total Language, Total Mathematics, Reference Skills, Science and Social Studies scores. Reading consists of Vocabulary and Comprehension; the Language area contains Mechanics, Expression, and Spelling; Mathematics consists of Computation, Concepts, and Applications tests; Reference Skills assesses knowledge of the use of a library, parts of books, and standard reference works; the Science test assesses the ability to classify objects, quantify data, recognize trends or valid hypotheses, analyze experimental design, and recall scientific facts; and Social Studies measures the ability to recall specific information, read maps and graphs, interpret verbal material, evaluate research designs, and employ logic in problem solving. Reliability coefficients yielded by the Kuder-Richardson Formula 20 ranged from .65 to .95 for all levels and all subtests. Validity information is discussed in the technical bulletin.

Content Evaluation Series; c1969-73; Grades 7-12; Houghton Mifflin Company.

The series consists of a group of end-of-year tests designed to assess student progress in specific subject areas.

Agribusiness Achievement Test by Roland L. Peterson, Leo M. Harvill, and James T. Horner; c1973; Grades 9-12.

Designed to measure mastery of high school agriculture programs, the test emphasizes both farming methods and the mechanical and managerial elements of modern agriculture. It consists of four subtests: Animal Science, Plant and Soil Science, Mechanics, and Management.

Language Arts Tests by Elsa Graser, Leonard Freyman, and Ruth Reeves; c1969; Grades 7-9.

The series consists of three tests: the Language Ability Test, the Composition Test, and the Literature Test. The Language Ability Test assesses the student's grasp of important principles underlying the construction of the English sentence and his ability to use sentence elements in standard sentence patterns. The Composition Test is an instrument for gauging progress toward writing proficiency. The Literature Test is designed to assess the student's ability to identify the principal types of literature and to differentiate among them; it also gauges the breadth and the depth of the student's feelings for literature. Split-half reliability coefficients corrected by the Spearman-Brown formula range from .92 to .95 for grades 7-9. Validity data are available in the technical manual.

Mathematics Test by Gilbert Ulmer; c1969; Grades 7-9.

This test is designed to measure student progress toward the currently accepted goals of the seventh; eighth; and ninth-grade mathematics program. Emphasis is placed on computational skills and mathematical concepts, and on the application of skills and concepts in problem solving. The test reflects the new approach while retaining the best of traditional mathematics. Split-half reliability coefficients corrected by the Spearman-Brown formula for grades 7-9 range from .79 to .90. Content and criterion-related validity information is available in the technical manual.

Modern Economics Test by Morris G. Sica, Sylvia Lane, and John D. Lafky; c1971; Grades 7-12.

The test was designed to measure individual comprehension of the essential concepts and principles of economics and their application to evaluating the functioning of the American economy, analyzing its persistent problems, and rationally judging the viability and validity of economic decisions implemented by decision makers. Content areas of the test include: The National Income, Monetary and Fiscal Policy, The Price System, and International Economics and Economic Development. The test yields a split-half reliability coefficient of .87. Content validity data is available.

Modern Geometry Test by Gerald S. Hanna; c1971; Grades 10-12.

Designed to assess the extent to which individuals and groups have mastered the content of high school geometry, the test items reflect the topic and process content of the typical geometry course based upon a study of current geometry textbooks, professional literature and authoritative recommendations. Non-parallel halves reliability corrected by the Spearman-Brown formula yielded a coefficient of .85. Content validity information is available in the manual.

Science Tests by Ernestine O'Connell; c1969; Grades 8-9.

The science tests appraise the student's science literacy and reasoning ability. They were designed to measure the student's grasp of the everyday vocabulary of science and to assess his understanding of basic concepts in physical science and earth science. Subtests of the Physical Science Test include Chemistry and Physics. Subtests of the Earth Science Test include Geology, Astronomy, and Meteorology. Split-half reliability coefficients corrected by the Spearman-Brown formula range from .74 to .81 for both tests over grades 7-9. Validity information is available in the technical manual.

Cooperative Achievement Tests; c1962-65; Grades 9-14; Addison-Wesley Publishing Company, Inc.

Batteries of achievement tests are provided in a number of subject areas.

Cooperative English Tests

Designed to test students' mastery of basic English skills, the battery consists of two tests. Reading Comprehension includes two parts: vocabulary and reading passages of varied style and content. English Expression assesses effectiveness in conveying exact meaning and mechanics, including usage, spelling, punctuation, and capitalization. Three forms of the tests are provided at each of two levels.

Cooperative Literature Tests

These open-book tests aid in identifying and correcting difficulties arising from independent reading. Tests examining the following works are available: The Bridge of San Luis Rey, Great Expectations, Hamlet, Huckleberry Finn, Julius Caesar, Macbeth, The Merchant of Venice, Moby Dick, The Odyssey, Oedipus the King, The Old Man and the Sea, Our Town, Pride and Prejudice, Pygmalion, The Red Badge of Courage, The Return of the Native, The Scarlet Letter, Silas Marner, and A Tale of Two Cities.

Cooperative Mathematics Tests

This series of tests is designed to measure achievement in the major content areas of mathematics. The following tests are available: Arithmetic, Structure of the Number System, Algebra I, Algebra II, Geometry, Trigonometry, Algebra III, Analytic Geometry, and Calculus.

Cooperative Science Tests

The Cooperative Science Tests measure achievement in four major areas of the junior-senior high school science curriculum: general science, biology, chemistry, and physics. Achievement is assessed in terms of knowledge and understanding of basic concepts and principles, ability to apply knowledge in problem situations, and ability to analyze and evaluate scientific ideas and procedures.

Advanced General Science Test

This is a more intensive and comprehensive treatment of the content at a somewhat higher level of difficulty than the General Science Test. Content areas include: astronomy, geology, meteorology, biology, chemistry, and physics. The test is appropriate for most

ninth-grade general science courses and for students of average or higher ability who are completing general science at the eighth-grade level. The Kuder-Richardson Formula 20 yielded a reliability coefficient of .94. Content validity has been examined.

Biology Test

This test is designed to measure the outcomes of a typical high school course in biology. Part I deals with human biology and the general structure, characteristics, and processes of living things. Part II is devoted primarily to plant and animal forms. Content areas include: The Nature of Life and of Science, The Cell, Characteristics of Life, Heredity and Change, Anatomy and Physiology, Nutrition, Hygiene and Disease, Major Plant Groups, Major Animal Groups, Ecology and Conservation, Distinctive Characteristics of Typical Forms, Anatomy of Typical Forms, Life Cycles of Typical Organs, and Physiology of Typical Organs. Reliability coefficients computed by the Kuder-Richardson Formula 20 ranged from .82 to .92. Content validity has been examined.

Chemistry Test

In this survey test of basic concepts and principles of chemistry, subject areas of Part I include: Matter, Reactions, Solutions, Structure, Electrical Relations, Elements, and Compounds. Part II includes: Laboratory Materials, Scientific Methods, Laboratory Skills and Techniques, Illustration and Clarification of Scientific Principles, and Laboratory Records and Reports. Reliability coefficients computed by the Kuder-Richardson Formula 20 ranged from .83 to .91 for grades 10-12. Content validity has been examined.

General Science Test

The General Science Test is a brief survey of elementary concepts in biology, chemistry, physics, astronomy, geology, and meteorology. It is appropriate for most introductory courses in grades 7, 8, or 9. Reliability computed using the Kuder-Richardson Formula 20 ranged from .89 to .92 for grades 7, 8, and 9, Forms A and B. Content validity has been examined.

Physics Test

This test is a survey of general concepts and principles of physics. Content areas of Part I include: Mechanics, Heat, Sound, Light, Electricity and Magnetism, and Modern Atomic Physics. Part II includes: Laboratory Materials, Scientific Methods, Laboratory Skills and Techniques, Illustration and Clarification of Scientific Principles, and Laboratory Records and Reports. Reliability coefficients computed by the Kuder-Richardson Formula 20 ranged from .82 to .91. Content validity has been examined.

Cooperative Social Studies Tests

The Cooperative Social Studies Tests measure achievement in seven major areas of the junior-senior high school social studies curriculum: American history (junior and senior high school), civics, world history, modern European history, American government, and problems of democracy. Achievement is assessed in terms of knowledge and understanding of events, issues, and ideas; application of knowledge in generalizing principles and concepts; analysis and evaluation of cause and effect relationships; and ability to use the accessories of the social sciences - graphs, maps, charts, cartoons, and so on.

American Government Test

The test is a measure of the important concepts and principles of the American Government. Subject matter includes: The Constitution and the National Government, State and Local Government, Citizenship and Political Participation, Government Services, Controls and Finances, and National Defense and International Relations. Reliability computed by the K-R Formula 20 yielded a coefficient of .90. Content validity information is available.

American History Test

The test is available on two levels, junior and senior high school. The general content of the two tests is similar, with both tests covering the period from exploration to the present. Subject matter includes: Relations with other Nations and Peoples, Government and Politics, Economic Development, and Social and Cultural Developments. The senior high school test places more emphasis on the last hundred years of United States history. Reliability coefficients computed using the Kuder-Richardson Formula 20 were .88 and .90 for junior and senior high school respectively. Validity information is available.

Civics Test

The Civics Test is an achievement test that places its emphasis on citizenship, political participation, government organization, services at all levels, and the constitutional basis for American government. Skills measured by the instrument include memory, understanding, and analysis of the Constitution and the National Government, State and Local Government, Citizenship and Political Participation, Government Services, Controls, and Finances, and National Defense and International Relations. A reliability coefficient of .90 was computed using the K-R Formula 20. Content validity has been examined.

Modern European History Test

This test is limited in its content scope to post-1450 Europe and her relationships with the rest of the world. Subject matter includes: Social and Cultural Developments, Political Developments, Economic Developments, and European Diplomatic Developments. A reliability coefficient of .87 was computed using the K-R Formula 20. Content validity has been examined.

Problems of Democracy Test

This device is designed to assess the important concepts, basic principles, and issues of democracy. Subject matter includes: International Affairs and National Defense, Political Problems, Economic Problems, and Social Problems. Reliability computed by the K-R Formula 20 yielded a coefficient of .90. Validity information is available in the handbook.

World History Test

Designed to assess knowledge of western civilization from ancient days to the present, the instrument also measures the historical developments in eastern Europe, Africa, and Asia. A reliability coefficient of .91 was computed using the K-R 20 Formula. Validity information is available.

Tests that are also available as part of the Cooperative Achievement series include:

- Cooperative French Listening Comprehension Test
- Cooperative Latin Test
- Cooperative Industrial Arts Tests
- AAHPER Cooperative Physical Education Tests

Cooperative Preschool Inventory: Revised Edition by Bettye M. Caldwell; c1967-74; Ages 3-6; Addison-Wesley Publishing Company, Inc.

This individually administered assessment and screening procedure is designed to provide a measure of achievement in areas regarded as necessary for success in school. The 64 items elicit information in the following areas: knowledge of self, ability to follow directions, verbal expression,

basic numerical concepts, and sensory attributes. The test yields scores for: Personal-Social Responsiveness, Associative Vocabulary, Concept Activation-Numerical, and Concept Activation-Sensory. A Spanish edition of the test is available. Kuder-Richardson 20 reliability coefficients range from .86 to .92 across the various age groups and corrected split-half reliability coefficients range from .84 to .93.

Cooperative Primary Tests; c1967; Grades 1-3; Addison-Wesley Publishing Company, Inc.

These tests are designed to measure skills and concepts that are basic to future success in reading, writing, listening, and mathematics. They can be used for diagnosis, program evaluation, academic assessment, or placement. It is suggested that each test user make an individual judgment of content validity with respect to his own instructional practices and aims. Two types of reliability were computed for each test. The battery consists of six tests:

Pilot

The 10-item test is designed to give children practice with the format and the kinds of questions and responses they will encounter in the tests. It can be helpful in determining if the children are ready to take the tests.

Listening

The test measures the ability to identify, recall, interpret, and draw inferences from material read aloud by the teacher. The child demonstrates his comprehension by marking appropriate pictures. Kuder-Richardson 20 reliability coefficients ranged from .79 to .86. Alternate form correlation coefficients were .75 and .82.

Mathematics

The test measures major mathematical concepts as they are developing in the young child. These concepts are arranged under the headings of number, symbolism, operation, function and relation, approximation and estimation, proof, measurement, and geometry. K-R 20 reliability coefficients ranged from .81 to .89. Alternate form correlation coefficients were .77 and .84.

Reading

The test assesses the child's ability to read and understand printed material. The vocabulary is geared to standard primary reading programs and is appropriately below that of the Listening tests. K-R 20 reliability coefficients ranged from .86 to .91. Alternate form correlation coefficients were .82 and .85.

Word Analysis

The test measures the child's understanding of phonetic and structural properties of words, emphasizing single vowels and consonants, blends, digraphs, and diphthongs at the beginning, middle, and end of words. K-R 20 reliability coefficients ranged from .88 to .93. Alternate form correlation coefficients were .82 and .92.

Writing Skills

Intended for administration in grades 2 and 3, the test assesses the child's ability to recognize correct spelling, punctuation and capitalization, and usage. K-R 20 reliability coefficients ranged from .80 to .85. Alternate form correlation coefficients were .80 and .88.

Coopersmith Self-Esteem Inventory (SEI) by Stanley Coopersmith; Not Dated; Ages 9-Adults; Self Esteem Institute.

There are two forms of the SEI. Form A provides a general assessment of self-esteem which may be broken down into component subscales depending on the goals and interests of the tester but which may also be used without such differentiation; Form B is briefer and does not permit further differentiation. There are five subscales which cycle in sequence the length of the SEI. These subscales are: General Self, Social Self-peers, Home-parents, Lie Scale, and School-academic. The total scores of Forms A and E correlate .86. Validity information is available.

Cross-Cultural Attitude Inventory by Steve Jackson and Rox Klinger; Not Dated; Ages 3-12; Dissemination Center for Bilingual and Bicultural Education.

This test is designed to measure the attitude toward the Mexican and Anglo cultures, and provides a score based on 11 items representative of each culture. The items chosen deal with language, facial characteristics, foods, games, clothing, sports, and flags, all of which are factors that make up the cultures. The total score for all 11 items gives a measure of how favorably the child perceives each culture, since the symbols cause the child to think of all the things he associates with these symbols. Validity information is available.

Cultural Attitude Inventories by Stephan L. Jackson; c1974; Grades 4-Adult; Learning Concepts.

These devices are intended for use in exploring the effects of programs designed to enhance ethnic identity or cross-cultural understanding and appreciation. Each of the two inventories contains 25 items representing the people, food, games, sports, language, art, music, and occupations of the Mexican-American and Anglo-American cultures. The student indicates his feelings toward and familiarity with each item. Both inventories are completed by the examinee who should be able to read the English included in the stimuli and response modes. The inventories are not designed for individual diagnosis. Normative tables are not included in the manual and the author recommends that each program use its own local control group for comparison purposes. Item-test correlations range from .41 to .73. Alpha reliability coefficients for Feelings and Familiarity are above .90 for both inventories.

Cultural Attitude Scales by Perry A. Zirkel and Stephan L. Jackson; c1974; Ages 5-14 Years; Learning Concepts.

The scales represent a modular approach to the measurement of cultural attitudes toward and knowledge of Puerto Rican, Anglo-American, Black-American, or Mexican-American cultures. They are applicable to programs intended to enhance ethnic identity or cross-cultural understanding among any one or more of these four ethnic groups. The instruments do not require reading ability since they use pictorial stimuli and response options. Directions are provided in both English and Spanish. Each scale yields two scores: a cultural attitude index and a cultural knowledge index. Split-half reliability coefficients adjusted by the Spearman-Brown formula are .68 for the Black-American scale, and .77 for the Puerto Rican and Anglo-American scales. Test-retest reliability coefficients obtained over a three-week interval ranged from .52 to .61. Evidence for content, construct, and criterion validity is presented in the manual. Technical data are not yet available for the Mexican-American scale.

Davis Reading Test by Frederick B. Davis and Charlotte C. Davis; c1956-62; Grades 8-13; The Psychological Corporation.

The test is designed to assess the overall reading ability of individuals or groups. Five categories of reading skills are required: finding the answers to questions answered explicitly or in paraphrase in a passage, weaving together the ideas in a passage and grasping its central thought; making inferences about the content of a passage and about

the purpose or point of view of its author; recognizing the tone and mood of a passage and the literary devices used by its author; and following the structure of a passage. Parallel forms reliability coefficients range from .74 to .91 for grades 8-13. Content and predictive validity have been examined.

Defining Issues Test by James Rest; c1972; Grades 9-16; James Rest

This instrument indicates an individual's appreciation of different conceptual stages in analyzing moral dilemmas by assessing the way he judges the importance of issues involved in a number of moral dilemmas. The examinee reads a story describing a moral dilemma and is then presented with 12 issues or considerations bearing upon that situation. The examinee evaluates each of the 12 issues and indicates its relative importance in deciding what ought to be done. Each issue was designed to exemplify some distinctive characteristic of a stage of moral development. The test is comprised of six stories and 72 issues. The test-retest reliability coefficient after two weeks is .81; after two years it is .57. Validity is examined extensively.

Diagnostic Mathematics Inventory (DMI) by John K. Gessel; c1975; Grades 1.5-8.5 and beyond for remedial students; CTB/McGraw-Hill.

This improved series of criterion-referenced tests measures student mastery of mathematics objectives. The DMI provides assessment from grades 1.5 to 8.5 in seven levels: A/Red, grades 1.5-2.5; B/Green, grades 2.5-3.5; C/Blue, grades 3.5-4.5; D/Orange, grades 4.5-5.5; E/Aqua, grades 5.5-6.5; F/Purple, grades 6.5-7.5; and G/Brown, grades 7.5-8.5.

In Level A/Red, counting and matching objectives precede those relating to addition and subtraction of whole numbers. Following are single-digit operations, multiple-digit addition, measurement, sequences, missing addends, geometry, and numeration systems.

For Level B/Green, counting and matching are continued, regrouping categories are introduced for addition and subtraction, and multiplication is added. Measurement is expanded and preoperational concepts in fractions are initiated.

For Level C/Blue, all the basic operations with whole numbers and their properties are included. Multiplication is expanded and division is introduced. Fractions, geometry, measurement, place value, and problem solving are also covered.

Level D/Orange proceeds through the basic operations, and the objectives for the commutative, associative, and distributive properties follow. Measurement and geometry are arranged in order of increasing complexity.

Level E/Aqua proceeds through the basic operations and on into number theory, measurement, place value, geometry, and problem solving.

Level F/Purple includes topics involving simple operations with negative integers, sets, percent, and statistics.

Level G/Brown places emphasis on computation involving negative integers, mathematical sentences, operations with sets, complex geometric objectives, percent, graphs, metric geometry, precision of measurement, statistics, probability, and trigonometry.

Diagnostic Reading Scales: Revised Edition by George D. Spache; c1963-72; Grades 1-8 and senior high school levels with reading disability; CTB/McGraw-Hill.

This series of integrated tests is designed to provide standardized evaluations of oral and silent reading skills and of auditory comprehension. The tests may be used to determine the proficiency of normal and retarded readers at elementary school levels and of retarded readers at junior and senior high school levels. The scales are comprised of three Word Recognition Lists, 22 Reading Passages, and eight supplementary Phonics Tests. The instrument yields three reading levels for each student: an Instructional Level determined by oral errors and comprehension, an Independent Level found by testing silent comprehension, and a Potential Level measured by auditory comprehension. Reliability coefficients range from .87 to .96 for the Word Recognition Lists. Reliability computed by test-retest yields coefficients ranging from .84 to .88 for Instructional and Independent Levels. Reliability coefficients of stability for a five-month interval for Potential Level were .94, .67, and .90 for grades 1, 2, and 3 respectively. Content, construct, concurrent, and predictive validity have been assessed.

Diagnostic Tests and Self Helps in Arithmetic by Leo J. Brueckner; c1955; Grades 3-12; CTB/McGraw-Hill.

These tests are designed to enable teachers to make a developmental diagnosis of the skills of students in all operations with whole numbers, common fractions, decimal fractions, percents, and measures.

The instrument consists of three closely integrated groups of materials that are essential for an effective program of instruction in arithmetic. The groups include four screening tests, 23 analytical Diagnostic Tests, and 23 Self-Helps closely geared to the Diagnostic Tests. Results of these tests identify specific weaknesses in the work of students and the points at which the major skills break down. Reliability and validity information is available.

Dunning-Abeles Physics Test by Gordon M. Dunning and Sigmund Abeles; c1967; Grades 10-13; The Psychological Corporation.

Designed to measure the extent to which important educational objectives have been attained by students in typical high school physics courses, the test assesses knowledge, comprehension, and application of the principles of physics. Content areas include: Mechanics (kinematics, dynamics, energy), Electricity and Magnetism (static electricity, current electricity, magnetism), Atomic and Nuclear Physics (particle nature of light, atomic models, spectra, nuclear physics), Wave Motion and Light (periodic wave motion, light) and Kinetic-Molecular Theory (kinetic theory, heat and temperature). Split-half reliability coefficients corrected by the Spearman-Brown Prophecy Formula ranged from .78 to .91. Content validity is discussed in the manual.

Durrell Analysis of Reading Difficulty: New Edition by Donald D. Durrell; c1955; Grades 1-6; The Psychological Corporation.

This individually administered diagnostic measure consists of a series of tests and situations in which the examiner may observe in detail various aspects of a child's reading. It yields a detailed analysis of the following phases of reading difficulty: silent and oral reading, listening comprehension, word analysis, phonetics, faulty pronunciation, writing, and spelling. The materials consist of the spiral-bound Reading Paragraphs, a quick-exposure device with accompanying cards, and an Individual Record Booklet for recording results systematically. The manual suggests general plans for remedial teaching. Psychometric data are not provided.

Durrell Listening-Reading Series by Donald D. Durrell, Mary T. Hayes, and Mary B. Brassard; c1968-70; Grades 1-9; The Psychological Corporation.

This instrument is designed to provide a comparison of children's reading and listening abilities. Its purposes are to identify children with reading disabilities, and to measure the degree of retardation in reading as compared to listening. The series consists of tests in both listening and reading at three levels: Primary (grades 1-3.5); Intermediate (grades 3.5-6); and Advanced (grades 7-9).

The Primary Level yields scores on: Vocabulary Listening, Sentence Listening, Vocabulary Reading, and Sentence Reading. K-R 21 reliability coefficients range from .70 to .96 for all subtests.

The Intermediate Level subtests include: Vocabulary Listening, Paragraph Listening, Vocabulary Reading, and Paragraph Reading. Split-half, odd-even reliability coefficients corrected by the Spearman-Brown Prophecy Formula range from .85 to .96. K-R 21 coefficients range from .81 to .96.

The Advanced Level consists of the following: Vocabulary Listening, Paragraph Listening, Vocabulary Reading, and Paragraph Reading. Split-half, odd-even reliability coefficients corrected by the Spearman-Brown Prophecy Formula range from .81 to .98. K-R 21 coefficients range from .80 to .98. Validity information for all levels is presented in the manuals.

Durrell-Sullivan Reading Capacity and Achievement Test by Donald D. Durrell and Helen B. Sullivan; c1937-44; Grades 2.5-6.9; The Psychological Corporation.

The instrument has two tests and two levels, primary and intermediate. Subtests of the Reading Capacity Test designed to measure comprehension of spoken language include: Word Meaning and Paragraph Meaning. The Reading Achievement Test consists of four subtests, as follows: Word Meaning, Paragraph Meaning, Spelling, and Written Recall. Split-half reliability coefficients corrected by the Spearman-Brown Formula ranged from .78 to .96 for both tests, grades 3-6.

Education Apperception Test by Jack M. Thompson and Robert A. Sones; c1973; Preschool-Grade 9; Western Psychological Services.

This projective technique was developed to assess a child's perception of school and the educative process. The test

consists of 18 pictures depicting children in school-related situations. The pictures are intended to evoke fantasy in four major areas: reaction toward authority, reaction toward learning, peer relationships, and home attitude toward school. The test should be administered and interpreted by experienced school psychologists. Sample case studies are provided in the manual. No technical information is available; however, research is in progress.

Educational Goal Attainment Tests (EGAT) by Bruce W. Tuckman and Alberto P.S. Montare; c1975; Grades 7-12; Phi Delta Kappa, Inc.

This collection of tests was developed under contract from the Puerto Rican Congress of New Jersey and provides for assessment of a broad range of educational goals common to many communities. EGAT covers both cognitive and affective goals, consisting of knowledge, attitude, and behavior items. It is designed for school or district needs assessment and not for diagnosis of individual student performance or individual professional accountability. Both Spanish and English versions of the tests are available. Split-half reliability estimates (corrected by Spearman-Brown formula) were computed for each of the educational goal attainment test scores. The coefficients range from a low of .393 to a high of .950. The tests constituting this assessment device are:

Civics

The test assesses Knowledge of Citizenship (knowledge of civic rights, knowledge of civic responsibilities), Attitudes toward Citizenship (attitudes toward productive citizenship, attitudes reflecting respect for property), Knowledge of Democratic Ideals (knowledge of rights and responsibilities in a democracy, knowledge of the American heritage), and Behaving in a Democratic Way (behavior representing loyalty to country, behavior representing patriotism).

Human Relations

The test measures: Knowledge of Individuals and Cultures (knowledge of other people and cultures, knowledge of group process), Attitudes Reflecting Tolerance toward Individuals and Cultures, Attitudes toward Human Relations, Behavior Reflecting Positive Human Relations, Knowledge Related to Change, Attitudes Related to Change (attitudes reflecting tolerance of ambiguity, attitudes reflecting belief in internal control), and Non-Conformity Behavior.

English Language

The test assesses Language Arts Skills, including those pertaining to word meaning, language mechanics, and reading comprehension.

Reasoning

The test assesses abilities in Logical and Scientific Reasoning, including scientific methodology skills and problem-solving skills.

General Knowledge

The test assesses General Knowledge (knowledge of math and science, knowledge of current and historical events), and Knowledge of Mainstream Culture (knowledge of family culture, general knowledge of culture).

Self

The test assesses: Attitudes towards Self (attitudes towards one's achievements, attitudes of self-awareness, attitudes of self-worth), Attitudes of Good Values and Character (attitudes of moral values, attitudes of character), Behavior Reflecting Good Values and Character (ethical behavior, self-discipline behavior), Attitudes toward Self-Improvement (attitudes of curiosity, attitudes toward learning), and Behavior Reflecting Self-Improvement.

Careers

The test assesses Specific Job Knowledge (trade knowledge, business knowledge), Attitudes toward Good Workmanship, Job-Related Behavior, Knowledge Related to Career Awareness (knowledge of career information sources, knowledge of job opportunities), Attitudes Reflecting Career Maturity, and Behaviors Reflecting Career and Self-Awareness.

Arts and Leisure

The test assesses: Knowledge Related to Hobbies and Leisure Activities, Attitudes Related to Hobbies and Leisure Activities (attitudes toward outdoor pastimes, attitudes toward indoor pastimes), Behavior Related to Hobbies and Leisure Activities, Aesthetic Knowledge (knowledge of music, knowledge of art and literature), Attitudes Towards the Value of Aesthetics, and Behavior Reflecting Creative Expression.

Life Skills

The test assesses: Knowledge of Family Management Principles, Attitudes toward Family Management, Behavior Reflecting Family Management, Knowledge of Resource Management (knowledge of personal economics, knowledge of environment), Behavior Reflecting Resource Management, Knowledge of Health Practices, Attitudes toward Health Practices, Behavior Reflecting Health Practices (personal hygiene, exercise-diet-safety).

Latin America

The test assesses: knowledge of Puerto Rican Culture (knowledge of history in culture, general knowledge of culture of people-places-art), and Spanish Language Arts Skills (language mechanics, vocabulary and reading comprehension).

Ego-Ideal and Conscience Development Test by Russell N. Cassel;
c1969; Ages 12-18; Monitor.

This test is designed to assess the degree to which the ideas a person expresses as the ideal solutions to social problems agree with the ideal solutions expressed by the mainstream of society. For each of the 80 social problems presented, five multiple-choice answers are provided ranging from a positive and conforming response to a negative and totally nonconforming response. Social problems are presented for each of the eight areas of the ego-ideal: Home and Family, Inner Development, Community Relations, Rules and Law, School and Education, Romance and Psychosexual, Economic Sufficiency, and Self Actualization. Internal consistency reliability was computed for part and total scores using three modifications of the K-R Formula 20 and eight different groups. Face, construct, status, factorial, and predictive validity are discussed in the manual.

Family Adjustment Test by Gabriel Elias; c1952-54; Ages 12-Adults;
Psychometric Affiliates.

This test is designed to elicit information on the extent to which an individual's intrafamily relationships produce positive (homey) or negative (homeless) feelings. The instrument consists of 10 subtests: Attitudes toward Mother, Attitudes toward Father, Father-Mother Attitude Quotient, Oedipal, Struggle for Independence, Parent-Child Friction-Harmony, Interparental Friction-Harmony, Family Inferiority-Superiority, Rejection of Child, and Parental Qualities. The test can be administered orally to individuals or, as a written test to groups of any size. Split-half reliability of the test was .97 when corrected for attenuation. Validity studies are described in the manual.

Geist Picture Interest Inventory by Harold Geist; c1964-68; Grades 8-12 and Adults; Western Psychological Services.

This nonverbal inventory assesses quantitatively 11 male and 12 female general interest areas: Persuasive, Clerical, Mechanical, Musical, Scientific, Outdoor, Literary, Computational, Artistic, Social Service, Dramatic, and Personal Service (female only). It also aids in the identification of motivating forces behind occupational choice: Could Not Say, Family, Prestige, Financial, Intrinsic and Personality, Environmental, and Past Experience. The inventory is self-administering and can be used with individuals or groups. Test-retest reliability coefficients were computed on 15 samples for each of 11 interest areas. Data from empirical validation studies are also discussed.

Getting Along by Trudys Lawrence; c1964-65; Grades 7-9; Tests in Microfiche.

The test was developed to evaluate pupil behavior in everyday situations and to enable teachers to identify pupils who may need help in improving behavior. It is divided into three parts based on the analysis of emotional health: Part I - Getting Along with One's Self (self-acceptance); Part II - Getting Along with Others (acceptance by others); and Part III - Getting Along in One's Environment (facing reality). Reliability coefficients, based on Kuder-Richardson Formula 21, were computed by grade and age. The coefficients ranged from .57 to .86. Validity data are discussed in the teacher's manual.

Gilmore Oral Reading Test: New Edition by John V. Gilmore and Eunice C. Gilmore; c1968; Grades 1-8; The Psychological Corporation.

This test was developed to provide classroom teachers, reading supervisors, and others particularly concerned with reading instruction, with a means of analyzing oral reading performance. The test measures accuracy of oral reading, comprehension of material read, and rate of reading. The instrument serves as an aid in adjusting instruction in both oral and silent reading. Reliability coefficients range from .53 to .94 for accuracy, comprehension, and rate for grades 3 and 6. Validity information is discussed in the manual.

Gordon Occupational Checklist by Leonard V. Gordon; c1961; Grades 9-12; The Psychological Corporation.

Designed to provide an indication of an individual's occupational preferences, the Checklist consists of 240 activities that are performed in many different kinds of jobs. These activities are organized into five general interest categories: Business, Outdoor, Arts, Technology, and Service. Test-retest reliability coefficients for item response frequencies within the same sample were .83 for boys and .95 for girls; between two different samples the coefficients were .95 for boys and .93 for girls. Validity data are presented and discussed.

Graves Design Judgment Test by Maitland Graves; c1946-74; Grades 7-16 and Adults; The Psychological Corporation.

The test is designed to measure certain components of aptitude for the appreciation or production of art structure. The test accomplishes this measurement by evaluating the degree to which a subject perceives and responds to the basic principles of aesthetic order--unity, dominance, variety balance, continuity, symmetry, proportion, and rhythm. Reliability coefficients corrected by the Spearman-Brown formula range from .81 to .93. Validity information is provided in the manual.

Grieg Social Studies Test by Mary E. Grieg; c1964; Grades 6-8; Scholastic Testing Service, Inc.

Designed to assess students' factual knowledge of American history, civics, and geography as well as their ability to interpret historical information and to produce reasoned inferences based on their knowledge of historical processes, the 100 items constituting the test cover the following areas: knowledge of the functioning of the government of the United States; understanding of the relationship between time and events in U.S. history; knowledge of the important contributions of historical figures; ability to interpret historical information; understanding of the laws of the land; geographical concepts; and reasoned inferences. Split-half reliability coefficients computed for grades 6 through 8 ranged from .90 to .93. Kuder-Richardson coefficients ranged from .90 to .92 for the same grade range. Content and construct validity are discussed in the manual. Data on congruent and concurrent validity are being collected.

Hall Occupational Orientation Inventory: Second Edition by L.G. Hall; c1971; Grades 7-16 and Adults; Scholastic Testing Service, Inc.

The inventory provides a systematic framework designed to enable the individual to explore the factors involved in occupational choice in terms of the relationship between psychological needs and value fulfillment, worker traits, and job-content characteristics. It yields scores for 22 scales: Creativity, Risk, Information, Belongingness, Security, Aspiration, Esteem, Self-Actualization, Personal Satisfaction, Routine-Dependence, Data Orientation, Things Orientation, People Orientation, Location Concern, Aptitude Concern, Monetary Concern, Physical Abilities Concern, Environment Concern, Co-worker Concern, Qualifications Concern, Time Concern, and Defensiveness. The manual emphasizes that the inventory is intended to aid in career counseling and not to predict an occupation toward which the counselee is most oriented. Test-retest reliabilities were estimated for all of the scales and ranged from .74 to .93. Evidence for content, concurrent, and predictive validity is presented.

Harrison-Stroud Reading Readiness Profiles by M. Lucille Harrison and James B. Stroud; c1949-56; Kindergarten-Grade 1; Houghton Mifflin Company.

The test assesses specific abilities and skills that children use in beginning to learn to read. There are six subtests: Using Symbols, Making Visual Discriminations (attention span controlled and uncontrolled), Using the Context, Making Auditory Discriminations, Using Context and Auditory Clues, and Giving the Names of the Letters. The Profiles can be used for three purposes: to determine the level of development of specific readiness skills, to aid in placement, and to facilitate diagnosis. Aspects of validity are discussed in the manual.

Hartman Value Profile (HVP) by Robert S. Hartman and Mario C. Trigos; c1973; Ages 12 and Over; Research Concepts.

An axiological test that measures a person's capacity to value, this instrument also provides indications of the presence or absence of emotional and existential problems. Measures obtained from the instrument include: World Concept-Self Concept Potentials, Cognitive-Affective Domain Relationships, Social-Emotional Handicaps, and Interpersonal Compatibility. Interpretation of the test is made individually through the profile.

How I See Myself Scale by Ira J. Gordon; 1968; Grades 3-12; The Florida Educational Research and Development Council.

Designed as a means of estimating the child's views of self and school, the instrument is available in two forms; an elementary form and a secondary form. The scale yields several factor scores about the child's self-concept including the following: teacher-school, physical appearance, interpersonal adequacy, autonomy, and academic adequacy; secondary school factors also include physical adequacy, male factors only, and female factors only. Test-retest reliability yielded coefficients ranging from .62 to .89 for students and .45 to .82 for the mothers' form. Content and construct validity data are available.

Howell Geometry Test by Edgar N. Howell; c1969; Grades 9-12; The Psychological Corporation.

The test measures the extent to which important educational objectives have been attained by students in typical high school geometry courses. It assesses students' understanding, knowledge, and application of concepts and principles in the following areas: Points, Lines, Planes; Angles; Postulational Systems; Geometric Relations; Triangles; Quadilaterals and Other Polygons; Circles and Spheres; Construction; and Coordinate Geometry. Split-half reliability coefficients corrected by Spearman-Brown Prophecy Formula are .82 for the total norm group and range from .67 to .86 for various samples.

Individual Career Exploration: Experimental Edition by Anna Miller-Tiedeman; c1974; Grades 8-Junior College; Scholastic Testing Service, Inc.

This inventory is designed to help students focus on future occupations in relation to their current interests, experiences, abilities, and ambitions. Completion of the inventory enables students to narrow their occupational choices down to two occupational groups and two responsibility levels. They can then investigate specific occupations within these categories. The occupational groups are: Service, Business Contact, Organization, Technology, Outdoor, Science, General Culture, and Arts and Entertainment. The responsibility levels are:

Professional and Managerial-independent responsibility, Professional and Managerial, Semi-professional and Small Business, Skilled Labor, Semi-skilled Labor, and Unskilled Labor. The inventory is in five sections: Interests, Experience, Occupations, Abilities and Skills, and Responsibility Level. Technical information is not provided.

Inferred Self-Concept Scale: Experimental Form by Elizabeth L. Mc Daniel; c1969; Grades 1-6; San Felipe Press.

An experimental form used in individual evaluation and research, the scale was designed to appraise the self-concept of individuals from different classes and cultures. Reliability has been determined through examiner agreement, split-half method, inter-item consistency, and test-retest. Content, criterion-related and construct validity have also been examined.

Instructional Objectives Exchange; c1972-75; Kindergarten-Grade 12; The Instructional Objectives Exchange.

The IOX makes available two types of criterion-referenced material - Objectives-Based Tests and Objectives Collections.

Objectives-Based Tests

This series of criterion-referenced tests is designed for use in program evaluation and individual diagnosis. Each set consists of approximately 40 tests, each of which is keyed to explicitly formulated instructional objectives. The tests are distributed as boxed sets of preprinted spirit masters. Each master allows duplication of 200 to 300 copies. All of the tests are available in two comparable forms. Sets are provided in the following areas:

Language Arts:

Mechanics and Usage, K-6--Included in this set are tests of capitalization; punctuation; abbreviation; hyphenation; bibliographic form; envelope, letter, and invitation form; plurals; possessives; pronoun referents; degree forms; subject-verb agreement; irregular past participles; misplaced modifiers; and commonly confused words.

Word Forms and Syntax, K-6--This set covers parts of speech, verb tense and time, types of verbs, complete sentences, functions of parts of speech in sentences, types of subjects and direct objects, sentence patterns, sentence transforms, and types of clauses.

Composition, Library, and Literary Skills, K-6--This set includes sentence precision, outlining, paragraph development, paragraph transition, and types of paragraphs as well as alphabetization, dictionary use, fiction and non-fiction, the Dewey Decimal System, and card catalog use.

Mathematics:

Sets and Numbers, K-6--This set of tests covers sets, numbers, and rational numbers.

Operations and Properties, K-6--The set includes addition, subtraction, multiplication, division, and combined operations.

Numeration and Relations, K-6--The set contains tests on numeration, ratios and proportions, graphs, statistics and probability, and logic.

Measurement, K-6--Included in this set are tests of monetary measurement; linear measurement; liquid weight, distance, time, rate, area, and volume measurement; temperature measurement; pressure, density and concentration measurement; and scale drawings.

Geometry, K-6--This set contains tests regarding points, lines, planes, simple plane figures, curves, angles, parallelism, perpendicularity, triangle similarity and congruence, circles, segments, polygons, solids, constructions, formula use, and geometric symbols.

Elements, Symbolism, and Measurement, Grades 7-9--This set contains tests about sets, integers, rational numbers, real numbers, numeration, sentence and logic, and measurement.

Geometry, Operations, and Relations, Grades 7-9--Included in this test set are geometry (operations and properties of planes and solids), statistics, ratios and proportions, and graphs.

Reading:

Word Attack Skills, K-6--The set includes tests of visual and auditory discrimination, basic sight vocabulary, recognition of sounds and their association with letters, and pronunciation of patterned letter combinations and words.

Comprehension Skills, K-6-- This set includes tests of reading comprehension of main idea, conclusions, sequence, and context clues in text, as well as punctuation, syntactical structures, and affixes.

Social Studies:

American Government, Grades 10-12-- This set contains tests covering Introduction to Government, Our Colonial Heritage, the American Constitution, Government and the Individual, American Politics, the Congress, The Presidency and the Executive Branch, the Federal Judiciary, and State and Local Government.

Objectives Collections

This series of 45 separate collections of instructional objectives covers both cognitive and affective outcomes. Each of the Collections consists of a number of objectives, a pool of measurement items for each objective, and a means for assessing the adequacy of the student's responses to the items. A few of the Collections include:

Measures of Self-Concept: Revised

This is a collection of affective objectives dealing with four dimensions of the learner's self. These include: Family (self-esteem yielded from family interactions), Peer (self-esteem associated with peer relations), Scholastic (self-esteem derived from success or failure in scholastic endeavors), and General (comprehensive estimate of self-esteem). Self-report inventories, both direct and indirect, or observational indicators are provided to assess the attainment of each objective. The objectives and related assessment devices are arranged into three grade levels: primary, intermediate, and secondary.

Objectives Collection in American Government

The emphasis of this collection is on interpretation and analysis of data rather than on mere recognition of facts. The objectives and accompanying items are divided into nine major content areas: Introduction to Government, Our Colonial Heritage, American Constitution, Government and the Citizen, American Politics, Organization and Function of the Congress, Presidency and the Executive, Federal Judiciary, and State and Local Government.

Objectives Collection in Art

This collection is organized into the following sections: Elements of Art (color, texture, line, shape, space, value), Principles of Art (unity, rhythm, repetition, pattern, contrast, balance), Perceptual Skills, and Basic Techniques. Sample items, apart from identification items, are drawn from the areas of sculpting, print-making, painting, drawing, and constructing.

Objectives Collection in Attitude Toward School, Revised

This collection of affective objectives deals with the learner's self-concept. Dimensions employed are: Teachers (subjective feelings about teacher behavior), School Subjects (differential attitudes toward commonly taught school subjects), Learning (attitude toward the learning experience independent of attitude toward school, teachers, and subjects), Social Structure and Climate (attitude toward the school as a social center and a rule-making and enforcing agency), Peer (feelings about the structure and climate of relationships within the peer group), General (orientation toward schooling independent of a particular school). An observational indicator and both direct and inferential self-report measures are provided to assess the attainment of the objectives.

Objectives Collection in Attitudes Related to Tolerance

These affective objectives deal with tolerance for the values and opinions of others. Direct and inferential self-report measures, sociometric devices, and observational techniques are provided to assess attainment of the objectives. Attitudes toward diversity of experiences and ethnic groups are the bases for the measures.

Objectives Collection in Health (Nutrition)

These objectives concentrate on nutrition and are based upon four central concepts: foods from the four food groups supply the nutrients needed for growth and health; a balanced daily diet includes foods from each of the four food groups; man's environment influences his food choices; and through various food-processing methods, man has available a large variety of foods that are high in nutritional value and quality.

Objectives Collection in Judgment: Analyzing Fallacies and Weaknesses in Arguments

This collection covers: Informal Fallacies (fallacies of relevance, insufficient evidence, and ambiguity), Arguments that Compare Context (context mixing, metaphor, and analogies), and Soundness (valid and sound arguments).

Objectives Collection in Judgment: Deductive Logic and Assumption Recognition

These objectives relate to the student's ability to make judgments based upon logical analysis in both emotional and nonemotional situations. The logical skills assessed are in the areas of conditional reasoning, class reasoning, assumption recognition, and recognition of reliable observations.

Objectives Collection in Music

This collection reflects major concepts, fundamentals, and applications in music appreciation. The objectives are organized into the following categories: Melody; Harmony; Musical Form; Musical Sound; Musical Style; Rhythm Performance Skills; and Band Instruments.

Other Collections which are also available include:

Algebra, Grades 8-12
Anthropology, Grades 4-6
Auto Mechanics, Grades 10-12
Biology, Grades 10-12
Bookkeeping, Grades 10-12
Business, Grades 10-12
Composition Skills, K-12
Data Relationships, K-9
Decoding Skills of Reading, K-12
Early Childhood Education, Prekindergarten
Electronics, Grades 7-12
Environmental Education, Grades 4-9
Figure and Object Characteristics, K-9
French, Grades 7-12
General Business, Grades 10-12
General Metals, Grades 7-12
Geography, K-6
Geography, Grades 7-9
Geometry, Grades 10-12
Home Economics, Grades 7-9
Home Economics, Grades 10-12
Knowledge and Attitudes of Drug Use, Grades 4-12
Life Sciences, K-6
Literature, Grades 7-9
Literature, Grades 10-12
Listening, Oral Expression, and Journalism, K-12
Mechanical Drawing, Grades 7-12
Mechanics and Usage, K-12

Numbers and Their Operations, K-9
Physical Education, K-6
Physics, Grades 10-12
Reading Comprehension, K-12
Secretarial Skills, Grades 10-12
Sex Education, K-12
Sociology, Grades 4-9
Spanish, Grades 7-12
Structural Analysis, K-12
Study and Reference Skills, K-12
Symbolism and Notation, K-9
Traditional Grammar, K-12
Transformational Grammar, K-12
U.S. History, Grades 7-9
U.S. History, Grades 10-12
Woodworking, Grades 7-12

Inter-Person Perception Test by F.K. Heussenstamm and R. Hoepfner; c1969;
Ages 6-15+; Monitor.

This test is an attempt to assess people's nonverbal communication. Facial expressions, which have been shown to be highly culture specific, are used to measure person perception. Two interchangeable forms of the test are available, AC and AA. Although the reliability of the total Form AA or AC has not yet been determined empirically, the estimation for each test form is approximately .82. No validity studies have been performed as yet.

Iowa Silent Reading Tests: 1973 Edition by Roger Farr; c1973; Grades 6-16; The Psychological Corporation.

The tests survey the specific areas in which the individual must have developed competence in order to read effectively. These areas include vocabulary, reading comprehension, speed of reading with comprehension, and, at Levels 1 and 2, use of reference materials and skimming and scanning for specific information. There are three levels of the test covering grades 6 through 9, grades 9 through 14, and academically accelerated high school and college students. Overall, the internal consistency reliabilities range from .70 to .92. Correlations between Levels 1 and 2 of the test and the Metropolitan Achievement Test subtests of similar content are provided as evidence of concurrent validity.

Iowa Tests of Basic Skills, Primary Battery, Forms S and L by A.N. Hieronymus and E.F. Lindquist; c1972; Grades 1.7-3.5; Houghton Mifflin Company.

This is a comprehensive measurement of growth in the fundamental skills crucial in current daily learning activities and in future educational development. The battery is available in two forms (5 and 6) and two levels (7 and 8). Subtests of both levels include: Listening; Vocabulary; Word Analysis; Reading Comprehension, Pictures, Sentences, Stories; Language Skills, Spelling, Capitalization, Punctuation, Usage; Work-Study Skills, Maps, Graphs and Tables; Preferences; Mathematics Skills, Mathematics Concepts, and Mathematics Problems. The tests facilitate provision for individual differences in needs and abilities. Technical information is unavailable in the teacher's guide.

Iowa Tests of Music Literacy (ITML) by Edwin Gordon; c1970; Grades 4-12; Bureau of Educational Research and Services.

These tests are designed to sequentially assess fundamental musical achievement in tonal and rhythmic aural perception and in reading and writing tonal and rhythmic notation. The tests are comprised of six subtests that are classified into two divisions: Tonal Concepts and Rhythmic Concepts. Subtests of each division include: Aural Perception, Reading Recognition, and Notational Understanding. Reliability coefficients computed on a split-half basis and corrected through the use of the Spearman-Brown Prophecy Formula range from .71 to .92 for grades 4-12. Validity information is available in the manual.

JIM Scale by Jack R. Frymier; c1967; Grades 7-12; Tests in Microfiche.

This 80-item index was developed to provide a means for assessing students' motivation to learn in school. The split-half reliability coefficient is .70; test-retest reliability is .67. Extensive research has been conducted using the scale and results are discussed in the manual.

Jesness Behavior Checklist by Carl J. Jesness; c1970; Ages 10 and Over; Consulting Psychologists Press.

Designed to provide a systematic way of recording data about social behavior, the checklist consists of 30 items describing behavioral units encompassing a broad range of nonintellectual, noncognitive social behaviors. The items measure 14 bipolar behavioral factors: Unobtrusiveness vs. Obtrusiveness, Friendliness vs. Hostility, Responsibility vs. Irresponsibility, Considerateness vs. Inconsiderateness, Independence vs. Dependence, Rapport vs. Alienation, Enthusiasm vs. Depression, Sociability vs. Poor Peer Relations, Conformity vs. Nonconformity, Calmness vs. Anxiousness, Effective Communication vs. Inarticulateness, Insight vs. Unawareness and Indecisiveness, Social Control vs. Attention-Seeking, and Anger Control vs. Hypersensitivity. There are two forms of the checklist. On one, the individual is rated by an observer; the other involves self-appraisal. Test-retest reliability coefficients for the 14 scales ranged from .09 to .51. Corrected inter-rater reliability coefficients for the scales ranged from .63 to .80. The validity of the test is discussed in the manual.

Junior High School Test of Economics; c1974; Grades 7-9; Joint Council on Economic Education.

This test is designed to evaluate student understanding of basic economic concepts and principles and their use in dealing analytically with economic problems. The content categories of the test are: basic concepts (scarcity, opportunity costs, supply and demand); gross national product and its determinants; money, price, and inflation; government taxation and spending; economic growth; government policies; operation of a market economy; organization and role of the firm; factors of production and distribution of income; international trade; and comparison of economic systems. Kuder-Richardson 20 reliability coefficients for grades 7, 8, and 9 are .69, .76, and .78 respectively; KR 21 coefficients are .67, .74, and .76 respectively.

Key Math Diagnostic Arithmetic Test by Austin J. Connolly, William Nachtman, and E. Milo Pritchett; c1971; Kindergarten-Grade 7; American Guidance Service, Inc.

This individually administered test is designed to provide a diagnostic assessment of mathematics skills in three major areas: Content (numeration, fractions, geometry and symbols), Operations (addition, subtraction, multiplication, division, mental computation, numerical reasoning), and Applications (word problems, missing elements, money, measurement, time). Most items require the subject to respond verbally to open-ended items that are presented orally by the examiner. Illustrations generally accompany the oral presentation. The subtests with their respective items are organized sequentially. The diagnostic profile allows the simultaneous determination of a student's performance at the area level, subtest level, and item level, as well as his total test performance. Split-half analysis yielded reliability coefficients for total test performance by grade level ranging from .94 to .97. Evidence of face, content, and concurrent validities is discussed in the manual.

Knowledge of Occupations Test by LeRoy G. Baruth; c1974; Grades 9-12; Psychologists and Educators, Inc.

Designed to evaluate what students know about occupations, this test covers job descriptions, training, certification and licensing, trends, tools, terminology, earnings, and the ability to interpret occupational material. Results can be used in curriculum planning, instruction, and individual counseling. The Kuder-Richardson Formula 20 was used to calculate a reliability coefficient of .90. Content validity is discussed in the manual.

Kuder General Interest Survey by G. Frederic Kuder; c1963-64; Grades 6-12 and Adults; Science Research Associates, Inc.

This is a survey of interests in a wide range of activities. Interest areas measured include: Outdoor, Mechanical, Computational, Scientific, Persuasive, Artistic, Literary, Musical, Social Service, and Clerical. An interest profile is sketched indicating the high and low interest areas of the individual. The survey enables the student to select appropriate educational or vocational goals. Test-retest and K-R 20 reliabilities range from .70 to .92.

Kwalwasser Music Talent Test by Jacob Kwalwasser; c1953; Grades 4-16; Mills Music, Inc.

This instrument is designed to measure musical aptitude. Form A consists of 50 three-tone patterns which are repeated with variation in one of the following respects: pitch, time, rhythm, and loudness. Form B consists of 40 items that present the same features in a somewhat simplified form. Technical information is unavailable.

Lankton First-Year Algebra Test: Revised Edition by Robert S. Lankton; c1965; Grades 8-12; The Psychological Corporation.

This test measures the extent to which important objectives have been attained by students in first-year high school algebra. These objectives include a knowledge and application of the fundamental skills of algebra and an understanding of basic concepts including: Number, Set, Operation, Structure, and Relation. Split-half reliability coefficients for a variety of samples range from .65 to .88. Content or curricular validity is claimed.

Lee-Clark Reading Readiness Test by J. Murray Lee and Willis W. Clark; c1962; Kindergarten-Grade 1; CTB/McGraw-Hill.

This test is designed to provide the teacher with an objective basis for identifying children who are ready to receive reading instruction. The primary objective of each of the subtests is to provide information that will enable the teacher to predict a child's ability to read. They are: Letter Symbols, Concepts, and Word Symbols. Split-half reliability coefficient corrected for length using the Spearman-Brown formula was .96. Kuder-Richardson formula 21 reliability coefficient for total test score was .88. Validity data is discussed in the manual.

Lowry-Lucier Reasoning Test Combination by Ellsworth Lowry and Omer Lucier; c1956-66; Grades 5-16 and Adults; B.M. Farley.

This device is designed to estimate reasoning ability in relation to chronological age and academic potential when employed in conjunction with a standard intelligence test. It is also used in the selection of computer programmers or wherever logical thought process is a major factor in job performance. Two forms of the test exist, A and B. The LLR-A requires the subject to solve problems presented in verbal form; the LLR-B calls for the solution of problems in the form of spatial relations. Test-retest reliability ranges from .69 to .84 for grades 5, 6, and 7.

Meier Art Tests: Art Judgment by Norman C. Meier; c1929-42; Grades 7-12; Bureau of Educational Research and Services.

Designed to measure the ability to detect principle functions in art that make for greater aesthetic value, the test is as much a test of appreciation as it is a test of ability. Principles included are proportion, equilibrium, and rhythm. Split-half reliability corrected by the Spearman-Brown formula yielded coefficients ranging from .70 to .83. Validity is discussed in the examiner's manual.

Meier Art Tests II: Aesthetic Perception by Norman C. Meier; c1963; Grades 9-16 and Adults; Bureau of Educational Research and Services.

An assessment of the ability to identify aesthetically appealing work, the test measures the ability to detect whether the parts of a piece of art are organized in accordance with universal principles. The relationship between art principles as detected by the subject serves to measure aesthetic perception. Technical information is available.

Metropolitan Achievement Tests (MAT): 1970 Edition by Walter N. Durost, Harold H. Bixler, J. Wayne Wrightstone, George A. Prescott, and Irving H. Balow; c1970-71; Grades K.7-9.5; The Psychological Corporation.

This is a series of measures designed to inform teachers and administrators how much pupils have learned in important content and skill areas of the school curriculum. There are six levels of the MAT as follows: Primer, grades K.7-1.4; Primary I, grades 1.5-2.4; Primary II, grades 2.5-3.4; Elementary, grades 3.5-4.9; Intermediate, grades 5.0-6.9; and Advanced, grades 7.0-9.5.

The Primer Level subtests include: Word Analysis, Reading, and Mathematical Concepts. Split-half, odd-even reliability coefficients, corrected by the Spearman-Brown formula range from .89 to .92. Saupe's estimate of K-R 20 reliability yields coefficients ranging from .92 to .96.

Primary Level I subtests include: Word Knowledge, Word Analysis, Reading, and Mathematical Concepts. Split-half, odd-even reliability coefficients corrected by the Spearman-Brown formula range from .94 to .97. Saupe's estimate of K-R 20 reliability yields coefficients ranging from .88 to .96.

Primary Level II subtests include: Word Knowledge, Word Analysis, Reading, Spelling, Mathematical Computation, Mathematical Concepts, and Mathematical Problem Solving. Split-half, odd-even reliability coefficients corrected by the Spearman-Brown formula range from .89 to .96. Saupe's estimate of K-R 20 reliability yields coefficients ranging from .85 to .96.

Elementary Level subtests include: Word Knowledge, Reading, Language, Spelling, Mathematical Computation, Mathematical Concepts, and Mathematical Problem Solving.

Intermediate Level subtests include: Word Knowledge, Reading, Language, Spelling, Mathematical Computation, Mathematical Concepts, and Mathematical Problem Solving, Science, and Social Studies. Split-half, odd-even reliability coefficients corrected by the Spearman-Brown formula range from .88 to .96. Saupe's estimate of K-R 20 reliability yields coefficients ranging from .84 to .96.

Advanced Level subtests include: Word Knowledge, Reading, Language, Spelling, Mathematical Computation, Mathematical Concepts, Mathematical Problem Solving, Science, and Social Studies. Split-half, odd-even reliability coefficients corrected by the Spearman-Brown formula range from .91 to .97. Saupe's estimate of K-R 20 reliability yields coefficients ranging from .91 to .97. Since each school has its own curriculum, the content validity of the MAT must be evaluated by each school.

Metropolitan Readiness Tests: 1976 Edition by Joanne Nurss and Mary E. McGauvran; c1974-76; Kindergarten-Grade 1; The Psychological Corporation.

The tests were developed to provide measures of important prereading skills. Level I is intended for use from the beginning through the middle of kindergarten and consists of six subtests: Auditory Memory, Rhyming, Letter Recognition, Visual Matching, School Language and Listening, and Quantitative Language. Level II focuses on more advanced, higher level skills important in beginning reading and mathematics. Intended for use at the end of kindergarten through the beginning of grade one, it includes eight subtests: Beginning Consonants, Sound-Letter Correspondence, Visual Matching, Finding Patterns, School Language, Listening, Quantitative Concepts, and Quantitative Operations. An optional Copying test is provided for use with both levels. Kuder-Richardson Formula 20 reliability coefficients are .92 for Level I and .93 for Level II. A split-half correlation, corrected for length by using the Spearman-Brown formula is .93 for Level I and .94 for Level II. The content validity of the MRT is discussed and empirical evidence for its predictive validity is discussed in the Teacher's Manual Part II.

Minnesota Counseling Inventory by Ralph F. Berdie and Wilbur L. Layton; c1953-57; Grades 9-12; The Psychological Corporation.

This device is intended to elicit information concerning the personality dynamics, personality structure, and personality problems of young people. It yields scores for: Validity, Family Relationships, Social Relationships, Emotional Stability, Conformity, Adjustment to Reality, Mood, and Leadership. Some training in psychology and education is recommended for interpreting the inventory. Odd-even reliability, corrected by the Spearman-Brown formula, was computed for each of the scales on several samples. The coefficients ranged from .57 to .95. Test-retest reliability coefficients were also estimated and ranged from .56 to .93. Validation studies are described.

Minnesota Vocational Interest Inventory by Kenneth E. Clark and David P. Campbell; c1965; Grades 9-16 and Adults; The Psychological Corporation.

This inventory was developed to provide systematic information on the interest patterns of men in nonprofessional occupations. It is intended as an aid to counselors working with students and others who are contemplating occupations at the semi-skilled and skilled levels. For each person who completes the Inventory, scores are derived to provide an index of similarity between his interests and the interests of men in a variety of nonprofessional occupations. The inventory employs the forced-choice method. Scoring keys are available for 21 occupational scales and nine homogeneous scales. Test-retest reliability coefficients for the occupational and homogeneous scales range from .62 to .88. Extensive validity data are provided.

Mooney Problem Checklist: 1950 Revision by Ross L. Mooney and Leonard V. Gordon; c1950; Grades 7-16; The Psychological Corporation.

The checklist was designed to serve as an aid in eliciting information from students concerning their personal problems. There are three educational forms available covering different levels. The college and high school forms deal with 11 problem areas: Health and Physical Development; Finances, Living Conditions, and Employment; Social and Recreational Activities; Social-Psychological Relations; Personal-Psychological Relations; Courtship, Sex and Marriage; Home and Family; Morals and Religion; Adjustment to College (School) Work; The Future: Vocational and Educational; Curriculum and Teaching Procedure. The junior high school form covers only seven problem areas: Health and Physical Development; School; Home and Family; Money, Work, the Future; Boy and Girl Relations; Relations to People in General; and Self-Centered Concerns. The technical characteristics of the checklist are discussed in the manual.

Moral Judgment Scale by Karen A. Maitland and Jacquelin R. Goldman; circa 1974; Adolescents; Karen A. Maitland.

This instrument is based on Kohlberg's schema of moral development. It consists of 15 vignettes describing a major issue of moral judgment. Each of the vignettes is followed by a question aimed at evoking one particular issue of moral judgment and six alternatives representing

characteristic modes of thought about the issue at each of the six stages of moral development. Test-retest reliability coefficients were .83 for grades 7 to 12 and .60 for a grade 10 to 11 sample. The corrected split-half reliability was .71. Reliability computed by the Kuder-Richardson Formula 20 was .67. Validity data are not presented.

Murphy-Durrell Reading Readiness Analysis by Helen A. Murphy and Donald D. Durrell; c1964-65; Grades K.8-1.5; The Psychological Corporation.

This test assesses two abilities important to success in beginning reading: phoneme perception and familiarity with the names of capital and lower-case letters. A measure of learning rate for sight words is provided as a further aid in grouping children. The Analysis consists of three subtests: Phonemes, Letter Names, and Learning Rate. The corrected odd-even reliability coefficient is .98. Data describing the test's predictive validity are presented.

Music Achievement Tests (MAT) by Richard Caldwell; c1968-69; Grades 4-12; Follett Publishing Company.

The MAT is designed to provide an accurate measurement of achievement for some of the most important objectives of a music education program. It is appropriate for use with any music textbook series, providing a logical and continuing measurement program. The information offered by MAT is important to diagnostic work, program planning, curriculum revision, and evaluation of objectives. There are four tests to the series as follows:

Test 1 provides standardized and diagnostic data on three musical skills: pitch discrimination, interval discrimination, and meter discrimination. Reliability coefficients estimated by K-R 21 range from .83 to .92 for total scores, grades 4-12.

Test 2 provides data on three musical skills: major-minor mode discrimination, feeling for tonal center, and auditory-visual discrimination (pitch and rhythm), which is closely related to music reading. Reliability coefficients estimated by K-R 21 range from .79 to .96 for total scores, grades 4-12.

Test 3 provides information on four musical skills: tonal memory, melody recognition, pitch recognition, and instrument recognition. K-R 21 reliability coefficients range from .46 to .90 for total scores, grades 4-12.

Test 4 provides information on four musical skills: musical style, auditory-visual discrimination, chord recognition, and cadence recognition. K-R 21 reliability coefficients range from .81 to .88 for total scores, grades 5-12. Content and criterion-related validity information for all tests is discussed in the interpretive manuals.

Musical Aptitude Profile by Edwin Gordon; c1966; Grades 4-12; Houghton Mifflin Company.

This profile serves as an objective aid in the evaluation of students' musical aptitude in order that the teacher can better provide for individual needs and abilities. The test consists of three main parts: Tonal Imagery, Rhythm Imagery, and Musical Sensitivity. The first part has two sections, Melody and Harmony; the second part also has two sections, Tempo and Meter; the last part has three sections, Phrasing, Balance, and Style. Reliability estimates range from .80 to .90. Validity information is available.

My Self Checklist by Robert E. Valett; c1973; Grades 4-9; Fearon Publishers.

The major purpose of this checklist is to enable students to express their feelings about how they perceive themselves. The checklist consists of four sections. The first section is comprised of 40 self-descriptive items to which the students respond "yes", "no", or "unsure". The second section consists of 10 incomplete sentences which the students complete by writing in what they feel when they read each sentence. In the third section, the students are asked to list as many of their personal strengths and weaknesses as they can think of. They then indicate what they believe other people consider to be their strengths and weaknesses. In the final section, the students indicate any concerns or comments they may have about the checklist. A Spanish version of the checklist is also available.

NM Concepts of Ecology Test; c1973; Grades 6-12; Monitor.

This criterion-referenced test was designed to assess specific learner objectives in the area of ecology and conservation. It is available at two levels. Level One covers natural resources, pollution, plant/animal dependencies, life processes, natural balance, geographic evolution and conservation, and material adaptation. Part Two covers life processes, plant/animal dependencies, geographic evolution and conservation, soil conservation, and natural adaptation. The internal-consistency reliability of the test was estimated for sixth, ninth, and twelfth grades by computation of the alpha coefficient (.67, .67, and .74 respectively). Validity studies are in progress. Individual interpretations can be criterion- or norm-referenced. Percentile and stanine norms are provided.

NM Consumer Mathematics Test; c1973; Grades 9-12; Monitor.

This criterion-referenced test is designed to assess specific learner objectives in the area of mathematics ability in consumer applications. The test items require the student to solve consumer problems using the fundamental operations of addition, subtraction, multiplication, and division. The internal-consistency reliability of the test, estimated by computation of the alpha coefficient, was .62 for grade 9 and .75 for grade 12. Validity studies are in progress.

NM Consumer Rights and Responsibilities Test; c1973; Grades 9-12; Monitor.

This criterion-referenced test was designed to assess specific learner objectives in the area of rights and responsibilities that consumers have. These include: consumer protection laws, economic conditions and terms, insurance, purchase payment plans, personal finance, and product information. Internal-consistency reliability (alpha coefficient) based upon ninth- and twelfth-grade students was .65 and .71 respectively. Validity studies are in progress. Individual interpretation can be criterion referenced or norm referenced. Percentile and stanine norms are available.

National Achievement Tests: American History-Government-Problems of Democracy by Vincent McGarrett; c1942-66; Grades 8-14; Psychometric Affiliates.

The test is designed to assess students' understanding of American history, government, problems, and the democratic way of life. It also measures their ability to make sound inferences based on the relationships among historic personages and events. The test consists of five parts: Growth of an American Spirit, Growth of Democracy, The Constitution, Foreign Policy, Problems of American Democracy. A reliability coefficient of .94 was estimated by odd-even scores and the Spearman-Brown formula.

National Achievement Tests: Health Education by John J. Shaw and Maurice E. Troyer; c1946-64; Grades 7-13; Psychometric Affiliates.

This test is designed to provide a systematic measure of accepted health materials and practices. The test consists of 100 items. Sixty multiple-choice items deal with important health facts and concepts, while 40 true-false items apply health knowledge through student reactions to problem situations. The test yields scores for Nutrition; Mental Health; Health in Home, School and Community; and Safety and First Aid. Corrected split-half analysis yields a reliability coefficient of .92 for Form A.

National Achievement Tests: Health by Robert K. Speer and Samuel Smith; c1937-62; Grades 3-8; Psychometric Affiliates.

This test includes items that assess health knowledge. It consists of four parts: Recognizing Best Habits; Health Comparisons; Causes and Effects; and Health Facts. A reliability coefficient of .84 was obtained for the test.

Nelson Biology Test: Revised Edition by Clarence H. Nelson; c1965; Grades 9-13; The Psychological Corporation.

This test measures the extent to which important educational objectives have been attained by students in typical high school biology courses. The items assess students' knowledge, understanding and application of biological principles and concepts in the following areas: Living Things, Life Processes, Ecological Relationships, and Methodology and Research. Split-half reliability coefficient corrected by Spearman-Brown Prophecy Formula is .92 for the total norm group.

Nelson-Denny Reading Test: Forms C and D by James I. Brown, M.J. Nelson, and E.C. Denny; c1973; Grades 9-16 and Adults; Houghton Mifflin Company.

This test provides a measure of three major elements of reading ability: vocabulary, comprehension, and reading rate. It can be used to serve predictive, screening, and broadly diagnostic purposes. Test-retest reliability coefficients for total score range from .82 to .91 for four different grade levels.

Nelson Reading Test: Revised Edition by M.J. Nelson; c1962; Grades 3-9; Houghton Mifflin Company.

This measure of reading achievement consists of 175 items. One hundred items measure vocabulary and 75 measure reading comprehension. Test results can be used to compare the achievement level of a given pupil, class, or school with the national norms; to identify pupils who need additional work; or to provide the teacher with a means of comparing the pupil's level of reading achievement with his learning capacity. The alternate-forms method was used to derive reliability coefficients that ranged from a low of .88 to a high of .93 for grades 3 through 9. Validity data are presented in the manual.

New Mexico Career Education Test Series by Charles C. Healy and Stephen P. Klein; c1973; Grades 9-12; Monitor.

This battery of six criterion-referenced tests was designed to assess

specific learner objectives in the area of career education. Internal-consistency reliability coefficients are given for each of the tests and range from .51 to .87. Some validity data are provided. Individual interpretations of test scores can be criterion- or norm-referenced. Percentile and stanine norms are available for grades 9 and 12. The tests included in the series are:

NM Attitude Toward Work Test

deals with the student's appreciation of the personal and social significance of work including: positive attitude toward preparing for an occupation; feeling that work contributes to self-confidence, self-esteem, and self-actualization; belief that work leads to many benefits; acceptance of the desirability of the interdependence of people; and belief in the value of work for our society.

NM Career Development Test

assesses the student's understanding of what is involved in holding a job, including the feeling that success or failure is a function of his own actions rather than luck, knowledge of how to conduct himself properly on the job, and knowledge of the factors that influence advancement.

NM Career Oriented Activities Checklist

determines whether the student has taken important steps involved in making an occupational decision. These steps include: consulted sources of information about occupations, engaged in activities that would provide information about occupations the student is considering, obtained the high school training needed for the occupations being considered, and made definite post-graduation plans.

NM Career Planning Test

assesses the student's ability to make appropriate decisions about preparing for and selecting an occupation. This ability includes: knowledge of sources of information to consult to learn about various occupations, and knowledge of what actions should be taken in order to make a decision.

NM Job Application Procedures Test

determines whether a student knows how to apply for a job. The test covers: making inquiries, reading advertisements and using employment agencies; completing an application form; and knowing how to conduct oneself during an interview.

NM Knowledge of Occupations Test

assesses the student's understanding of the characteristics and requirements of specific occupations.

Ohio Vocational Interest Survey (OVIS) by Ayres G. D'Costa, David W. Winefordmer, John G. Odgers, and Paul B. Koons, Jr.; c1969-72; Grades 8-12; The Psychological Corporation.

Designed to assist students with their educational and vocational plans, OVIS combines a Student Information Questionnaire with the Interest Inventory to provide the student and his counselor with valuable background data for interpreting the interest scores. The 24 OVIS interest scales include: Manual Work; Machine Work; Personal Services; Caring for People or Animals; Clerical Work; Inspecting and Testing; Crafts and Precise Operations; Customer Services; Nursing and Related Technical Services; Skilled Personal Services; Training; Literary; Numerical; Appraisal; Agriculture; Applied Technology; Promotion and Communications; Management and Supervision; Artistic; Sales Representative; Music; Entertainment and Performing Arts; Teaching, Counseling and Social Work; and Medical. Test-retest reliability coefficients range from .72 to .90. Validity has also been investigated empirically.

Oral School Attitude Test by Juan Rivera; c1973; Kindergarten-Grade 3; Learning Concepts.

Designed to reflect a student's perception of the school environment, this test deals with three aspects of student attitude toward school: Interpersonal Relations, Student-Instruction Interaction, and General School Factor. The test consists of 29 oral stimuli, each designed to elicit a response from the student that will be representative of his educational experiences and environment. In response to each of the 29 items, the child marks one of a series of faces representing a range of feelings. The items can be administered in English or Spanish. The test-retest reliability coefficient obtained over ten days was .77.

Orleans-Hanna Algebra Prognosis Test by Joseph B. Orleans and Gerald S. Hanna; c1968-71; Grades 7-11; The Psychological Corporation.

Intended primarily for use during the term that precedes students' enrollment in first-year algebra, this test is designed to predict their achievement. The test items pertain to four variables: Ability, Proficiency, Interest, and Other Personality Characteristics. Test-retest reliability coefficient for total test score is .95. Validity data are presented and discussed in the manual.

Orleans-Hanna Geometry Prognosis Test by Joseph B. Orleans, and Gerald S. Hanna; c1968-71; Grades 8-11; The Psychological Corporation.

This test is designed to predict achievement in a beginning course in geometry. It is intended primarily for use during the term that precedes students' possible enrollment in a geometry course. The test items pertain to four factors related to students' success in learning geometry: Ability, Proficiency, Interest, and Other Personality Characteristics. Test-retest reliability coefficient obtained for the total test is .86. Validity of the test is discussed in the manual.

Parent-Adolescent Communication Inventory by Millard J. Bienvenu; c1968-69; Adolescents; Family Life Publications, Inc.

This device was designed to help counselors, educators, and researchers assess communication in the parent-adolescent relationship for the purpose of individual counseling. It also helps the adolescent to focus on his relationship with his parents in an effort to promote better understanding of his interaction with them. Using the Spearman-Brown formula, a split-half correlation coefficient of .86 was computed. Test-retest studies on two different samples yield reliability coefficients of .78 and .88. Validity of the Inventory is examined in the manual.

Personal Adjustment Inventory by Carl R. Rogers; c1961; Grades 9-13+; Association Press.

This instrument was devised to assess the extent to which a child is satisfactorily adjusted to his peers, his family, and himself. The inventory attempts to cover most of the areas of child life in which maladjustment is apt to occur. It is also intended to reveal the child's methods of meeting his difficulties. Test-retest reliability yielded a coefficient of .72. Validity information is available in the manual.

Personal Orientation Inventory by Everett L. Shostrom; c1962-74; Grades 9-16 and Adults; Educational and Industrial Testing Service.

The inventory was developed to provide a comprehensive measure of values and behavior seen to be important in the development of the self-actualizing person. It consists of 150 two-choice comparative value and behavior judgments. The items are scored twice, first for two basic scales of personal orientation, inner-directed support and time competence and second for ten subscales each of which measures an important element of self-actualizing. These subscales are: Self-Actualizing Value, Existentiality, Feeling Reactivity, Spontaneity, Self-Regard, Self-Acceptance, Nature of Man, Synergy, Acceptance of Aggression, and Capacity for Intimate Contact. Test-retest reliability coefficients for the 12 POI scales range from .52 to .82. Extensive validity data are presented in the manual.

Personal Values Abstract by Harrison G. Gough; c1956-72; Grades 9-16 and Adults; Consulting Psychologists Press, Inc.

A self-report inventory containing scales for modernity, socialization, and femininity. Items for the modernity scale taken from the California Personality Inventory are concerned with dominance, status, sociability, social presence, and self-acceptance. Socialization refers to responsibility, stability, tolerance, reliability, etc. The femininity scale attempts to define the meanings of masculinity and femininity. Technical information is unavailable.

Pictorial Study of Values by Charles N. Shooster; c1957; Grades 9-12 and Adults; Psychometric Affiliates.

This test provides a way to measure reactions to six basic value areas: aesthetic, economic, political, religious, social, and theoretical. The test items consist of pictorial stimuli representing value-related scenes. For each scene, the respondent is asked to indicate the degree to which he would like to engage in the activity depicted. The test can be used with illiterates, persons with minimal reading skills, and non-English-speaking persons. The manual describes research that shows that the instrument correlates as well with self-judgments of values as does the Vernon-Allport-Lindzey Study of Values. No research is presented concerning reliability.

Piers-Harris Children's Self Concept Scale by Ellen V. Piers and Dale B. Harris; c1969; Grades 3-12; Counselor Recordings and Tests.

A self-report instrument designed primarily for research on the development of children's self attitudes and correlates of these attitudes, the scale consists of items that reflect the concerns that children have about themselves. Categories included are: Physical Characteristics and Appearance; Clothing and Grooming; Health and Physical Soundness; Home and Family; Enjoyment of Recreation; Ability in Sports, Play; Ability in School, Attitudes Toward School, etc.; Intellectual Abilities; Special Talents (music, arts); Just Me, Myself; and Personality, Character, Inner Resources, and Emotional Tendencies. The K-R Formula 21 was employed with resulting coefficients ranging from .78 to .93. The Spearman-Brown odd-even formula was applied with resulting coefficients of .87 to .90. Test-retest reliability coefficients ranged from .71 to .77. Content and concurrent validity has been examined.

Planning Career Goals by American Institutes for Research; c1975-76; Grades 8-12; CTB/McGraw-Hill.

This battery of tests and materials is designed to assist guidance and counseling personnel in helping students make realistic educational and career plans. PCG consists of four types of measures:

Interest Inventory

The 300 items in this instrument are presented in three sections: Occupations, Occupational Activities, and Current Activities. The examinee indicates interest in each of 12 career groups by rating job titles, job activities, or job-related youth activities on a five-point verbal scale.

Information Measures

This test samples knowledge that individuals would have acquired if they had studied about an occupation or participated in activities related to an occupation.

Ability Measures

This test contains 366 items within 10 sections: Reading Comprehension, Mathematics, Abstract Reasoning, Creativity, Mechanical Reasoning, English, Quantitative Reasoning, Vocabulary, Visualization, and Computation.

Life and Career Plans

This survey determines the student's present education and career plans and the importance of various personal needs beyond occupational roles.

Prescriptive Mathematics Inventory (PMI) by John K. Gessel; c1971-72; Grades 4-8; CTB/McGraw-Hill.

A criterion-referenced test designed to provide evaluation relevant to classroom instruction, the PMI measures a student's attainment of specific learning objectives. It assesses each student's mathematical knowledge and diagnoses his specific weaknesses. The test items represent an inventory of specific mathematics objectives over four levels: Orange Book (Grades 4 and 5), Aqua Book (Grades 5 and 6), Purple Book (Grades 6 and 7), and Level C (Grades 7 and 8).

The Orange Book contains 102 items involving addition, subtraction, multiplication, division of whole numbers, properties of these operations, number theory, measurement, nonmetric geometry, place value, and problem solving.

The Aqua Book contains 136 items including the basic operations with whole numbers, fractions, and decimal numbers. The properties of these operations are also tested, as are number theory, measurement, nonmetric geometry, place value, and problem solving.

The PMI Purple Book contains 169 items including the basic operations with whole numbers, fractions, and decimal numbers. Other items deal with the properties of these operations, with number theory, measurement, nonmetric geometry, numeration systems, percent, sets, statistics, and problem solving.

Level C contains 193 items including operations with negative numbers, number theory, numeration systems, sets, nonmetric geometry, percent, functions, measurement, statistics and probability, trigonometry, and reasoning.

Prescriptive Reading Inventory (PRI); c1972-75; Grades 1.5-6.5; CTB/McGraw-Hill.

The PRI is designed to diagnose the reading behavior of individual students and to provide information that can be used to reinforce, remediate or supplement their reading development. The inventory is comprised of components in four overlapping levels: Red Book (Level A, Grades 1.5-2.5), Green Book (Level B, Grades 2.0-3.5), Blue Book (Level C, Grades 3.0-4.5), Orange Book (Level D, Grades 4.0-6.5). Each book tests a series of reading behaviors appropriate to a range of age and grade levels in the elementary curriculum. Behavioral objectives measured include: Recognition of Sound and Symbol, Phonic Analysis, Structural Analysis, Translation, Literal Comprehension, Interpretive Comprehension, and Critical Comprehension. The technical report is scheduled for publication in the future.

Primary Reading Profiles by James B. Stroud, Albert N. Hieronymus, and Paul McKee; c1953-68; Grades 1-3; Houghton Mifflin Company.

This diagnostic battery of five tests is designed to evaluate pupil progress in reading. The tests are available on two levels, Level One and Level Two. Level One evaluates pupils at the end of first year instruction; Level Two evaluates pupils at the end of second year instruction. Subtests of both levels include Aptitude for Reading, Auditory Association, Word Recognition, Word Attack, and Reading Comprehension. The split-half reliability coefficient corrected by the Spearman-Brown formula was .97 for both levels. Content and concurrent validity have been examined.

Primary Self-Concept Inventory by Douglas G. Muller and Robert Leonetti; c1973-74; Preschool-Grade 6; Learning Concepts.

This test is designed to provide a procedure for evaluating several aspects of self-concept relevant to school success. While the test was constructed specifically for use with the child of Spanish or Mexican descent in the Southwest, it is also appropriate for use with children of many other cultural backgrounds. The test measures six aspects of self-concept which are clustered into the three major domains of: personal-self, social-self, and intellectual-self.

Principles of Democracy Test by Nathaniel L. Gage, Neil F. Garvey, Charles B. Hagan, and Roland Payette; c1960-61; Grades 9-12; Science Research Associates, Inc.

The test is designed to measure student knowledge and understanding of democratic principles as they are interpreted in the U.S.A. It covers: Constitution and fundamental law of the U.S.; elections, voting, and ballots; Congress; the president and executive branch; the judiciary and law enforcement; powers of the national government; state and local government; Declaration of Independence and Articles of Confederation; and civil liberties and the Bill of Rights. The test was standardized on a representative sample of high school students. Odd-even reliabilities, corrected by Spearman-Brown formula, range from .85 to .89. Reliability coefficients estimated by KR-20 range from .85 to .89. Evidence of content and concurrent validity is discussed in the manual.

QUESTA; c1971-72; Grades 9-12; Addison-Wesley Publishing Company, Inc.

Part of the Secondary School Research Program, The Questionnaire for Students, Teachers, and Administrators (QUESTA) is intended to elicit information from students concerning their attitudes toward their educational experiences. QUESTA is in two parts. QUESTA I is administered to students when they enter high school. It gathers baseline biographical, socioeconomic, and demographic information and identifies the student's attitudes toward himself, his peers, his previous school, and his new school, and his hopes, fears, and aspirations. QUESTA II is administered after the student has been in the school for some time--in order to measure the impact of the school upon his attitudes and values. The items in QUESTA II deal with three aspects of the school environment: personal characteristics of the people in the environment, other characteristics of the school itself and people's interaction with each other, and the effects of the school experience on students.

Rating Scale of Vocational Values, Interests, and Aptitudes by George D. Demos and Bruce Grant; c1966; Grades 9-16; Educational and Industrial Testing Service.

Developed as an aid in career counseling, the scales were designed to integrate a student's self-assessment of vocational values, interests, and aptitudes for several occupational areas. Technical data are not available.

Remote Associates Test: High School Form by Sarhoff A. Mednick, and Martha T. Mednick; c1971; Grades 9-12; Houghton Mifflin Company.

This test attempts to measure an aspect of creative ability defined as "the forming of associative elements into new combinations". Each of the 40 test items consist of a set of three words drawn from mutually remote associative clusters. The respondent must find a fourth word which could serve as a specific kind of associative connecting link. The odd-even reliability coefficient as estimated by the Spearman-Brown formula is .90. Validity data are not provided for the high school form.

SCORE: School Curriculum Objective-Referenced Evaluation; c1973-present; Kindergarten-Grade 9; Westinghouse Learning Corporation.

A criterion-referenced testing system that combines behaviorally stated objectives with a computer-retrievable bank of related items, this system enables school districts to obtain tests individually tailored to their curricular needs. The objectives and items cover four subject areas at three levels: primary (K-3), intermediate (grades 4-6), and junior high school (grades 7-9).

Language Arts/Reading

This collection is presented in two parts: The Reading section is concerned with ability to understand the relationship between sound and symbol, read and comprehend words and passages, employ a variety of study skills, and classify and analyze forms of literature. The Language section deals with ability to comprehend speech, understand and use the parts of speech, and write both logically and creatively.

Mathematics

Objectives constituting this collection deal with ability to understand the use of sets, numbers, mathematical operations, fractions, units of measurement, metric and nonmetric geometry, functions and functional relationships, tables and graphs, probability, statistics, and logic.

Science

This collection of objectives covers the following: biological science; physical science; earth science and astronomy; laboratory skills, career education, and the nature of science; and processes of science.

Social Science

The objectives in this collection deal with such subjects as man's basic needs, the family unit, social groups, social problems, American and world history, economics, geography, government, political systems, sociology, and anthropology.

SRA Assessment Survey: Iowa Tests of Educational Development by E.F. Lindquist and Leonard S. Feldt; c1970-72; Grades 9-12; Science Research Associates, Inc.

This instrument is intended to provide measures of educational development that are appropriate for all high school students, regardless of the specific curriculum they are following. The tests include: Reading (Comprehension and Vocabulary), Language Arts (Language Usage and Spelling), Mathematics, Social Studies, Science, and Use of Sources. KR 20 reliabilities range from .82 to .94 for Form X-5, and .81 to .95 for Form Y-5. Content, criterion-related, and construct validity have been assessed.

SRA Assessment Survey: Multilevel Edition of the Achievement Series by Robert A. Naslund, Louis P. Thorpe, and D. Welty Lefever; c1971-72; Grades 4-9; Science Research Associates, Inc.

This measure of educational development in basic curriculum areas consists of three separate but overlapping tests of graduated difficulty. The blue level is the easiest, the red the most difficult. The green overlaps the upper end of the blue level and the lower end of the red level. Each color-coded booklet contains Reading (including Vocabulary), Mathematics, Language Arts (including Spelling), Social Studies, Use of Sources, and Science tests. Reliability estimates range from .82 to .96 for all three levels.

SRA Assessment Survey: Primary Edition of the Achievement Series by Robert A. Naslund, Louis P. Thorpe, and D. Welty Lefever; c1972; Grades 1-4; Science Research Associates, Inc.

Designed to measure the general objectives of instruction in the primary grades, the test is available in two levels: Primary I-grades 1-2, and Primary II-grades 2-3. Both levels explore the

three basic skill areas of Language Arts, Reading, and Mathematics. Each area is divided into specific skills: Language Arts includes Alphabetization, Capitalization, Punctuation, Spelling and Usage; Reading consists of Reading Comprehension and Vocabulary; Mathematics consists of Math Concepts and Math Computation. K-R 20 reliability estimates range from .72 to .94 for Primary I and II, Forms E and F.

SRA Reading Record, Second Edition by Guy T. Buswell; c1947-59; Grades 6-12; Science Research Associates, Inc.

This is a diagnostic device that pinpoints the students' strengths and weaknesses in reading skills. A wide variety of reading skills is measured with items covering 10 kinds of reading situations: Reading Rate, Comprehension, Paragraph Meaning, Directory Reading, Map-Table-Graph Reading, Advertisement Reading, Index Usage, Technical Vocabulary, Sentence Meaning, and General Vocabulary. A profile of these skills is provided for use by both teacher and student. Kuder-Richardson Formula 21 reliabilities range from .52 to .86. Spearman-Brown reliabilities range from .59 to .92. Content validity has been examined.

STS Junior Inventory by Hermann H. Remmers and Robert H. Bauernfeind; c1957-68; Grades 4-8; Scholastic Testing Service.

This instrument is a downward extension of the STS Youth Inventory.

STS Youth Inventory by Hermann H. Remmers and Benjamin Shimberg; c1956-71; Grades 7-12; Scholastic Testing Service.

This needs and problems checklist enables students to express the personal importance of their interests and problems and their concerns about the direction of their educational and social growth. The inventory consists of 167 statements arranged under five broad headings: My School, After High School, About Myself, Getting Along with Others, and Things in General. Split-half reliability coefficients were computed for the five area scores and ranged from .86 to .96. Additional technical information, including a discussion of validity, is available in the manual.

San Diego County Inventory of Reading Attitude; c1961; Grades 1-6;
San Diego County Department of Education.

This survey instrument is designed to provide teachers with systematic information on children's attitudes toward reading. A split-halves reliability coefficient of .79 was obtained for the inventory. Application of the Spearman-Brown Prophecy Formula resulted in a reliability coefficient of .89. Validity of the instrument is discussed in the manual.

Scale of Socio-Egocentrism: 1975 Revision by James S. Peters; c1964-75;
Grades 7-16; Peters and Associates.

The scale evaluates the degree of adjustment a person makes to "self" and to "society". If evaluation is negative or extreme in either direction, indicating either egocentrism or sociocentrism, then the potential for maladaptive adjustment and delinquency proneness is postulated. The 60 items fall into seven categories of attitudes and feelings: toward self, toward primary personal relationships, toward members of the peer group, toward members of the opposite sex, toward secondary relationships, toward impersonal relationships, and toward "generalized others". Technical information is not provided.

School Attitude Survey by Harold F. Burke; c1970-72; Grades 4-6; Arden Press.

This device attempts to probe children's attitudes pertaining to an academic setting. The survey is divided into four sections: feelings about classroom learning, feelings about the teacher, feelings about peers, and feelings about the classroom environment. Use of the instrument is discussed extensively in the manual. Technical data are not available.

School Interest Inventory by William C. Cottle; c1959-66; Grades 7-12;
Houghton Mifflin Company.

This 150-item self-report inventory is a tool to be used in the early detection of potential dropouts. The manual suggests that results be used in conjunction with grades, teacher judgment, and other information

in determining which students to give priority counseling. Scores are interpreted on an intra-institutional basis. Test-retest reliability coefficients were computed for a number of samples. The coefficients ranged from .78 to .92. Extensive validity data are provided in the manual.

The School Inventory by Hugh M. Bell; c1936-63; Grades 9-12; Consulting Psychologists Press.

The inventory attempts to quantitatively describe pupils' attitude toward their school and provide an indication of their level of adaptation to the school environment. It should not be administered until the examiner has established a feeling of cooperation with the students being tested. An odd-even reliability coefficient of .94 was computed using the Spearman-Brown formula. Validity studies are described in the manual.

School Motivation Analysis Test by Arthur B. Sweney, Raymond B. Cattell, and Samuel E. Krug; c1970-76; Ages 12-18; Institute for Personality and Ability Testing.

The test provides a measure of 10 dynamic structures significant to human motivation. These 10 categories consist of six drives (assertiveness, mating, fear, narcissism, pugnacity-sadism, protectiveness), and four sentiments (self-sentiment, superego, school, home). Four measurement devices are used to assess the motivation factors: Autism, Utilities, Information, and Word Association. Technical information is not provided.

Seashore Measures of Musical Talent, 1956 Revision by Carl E. Seashore, Don Lewis, and Joseph G. Saetveit; c1939-60; Grades 4-16 and Adults; The Psychological Corporation.

The measure is designed to assess musical aptitude for the purposes of educational and vocational counseling admission to music instruction in schools, and selection for membership in bands and other musical organizations. The instrument provides separate measures for six capacities: pitch, loudness, rhythm, time, timbre, and tonal memory. Kuder-Richardson formula 21 reliability coefficients range from .55 to .85 for all six capacities, grades 4-16. Validity is discussed in the manual.

Self-Concept Adjective Checklist by Alan J. Politte; c1971; Kindergarten-Grade 8; Psychologists and Educators Press.

The checklist was developed for use by guidance counselors, social workers, school psychologists, and related professionals in their evaluation of elementary schools. It consists of 114 adjectives categorized as Physical Traits, Social Values, Intellectual Abilities, and Miscellaneous (emotional feelings, group behaviors, habits). The student responds to each adjective by marking a blank labelled "I am", "I am not", or "I would like to be". The reliability coefficient was .83 for a test-retest format, with an interval of six months. Cross-validated studies are in progress.

Self-Concept and Motivation Inventory (SCAMIN) by George A. Farrah; c1968; Preschool-Grade 12; Person-O-Metrics.

This group-administered, diagnostic inventory was developed to measure academic self-concept. It assesses four factors of self-concept: Goal and Achievement Needs, Failure Avoidance, Role Expectations, and Self Adequacy. The inventory is available at four levels: Preschool/Kindergarten, Early Elementary, Later Elementary, and Secondary. Students respond to questions prefaced with "What face would you wear if ..." by selecting faces from a three- or five-face response scale. Pilot studies produced Spearman-Brown split-half (odd-even) reliabilities of .93 for the total Secondary Form and .82 for the Later Elementary Form.

Self-Directed Search by John L. Holland; c1970-71; Ages 15-Adults; Consulting Psychologists Press.

The instrument is a self-administering, self-scored, and self-interpreted vocational counseling tool. In the assessment booklet, separate sections on Occupational Daydreams, Activities, Competencies, Occupations, and Self-Estimates determine a person's resemblance to each of six occupational or personality types: Realistic, Investigative, Artistic, Social, Enterprising, and Conventional. The respondent scores his responses and uses the resulting three-letter summary code to search the occupational classification booklet. KR 20's for the six scales in the Search range from .67 to .94.

Self-Directed Search: Form E by John L. Holland; c1970-73; Grades 4 and Above; Consulting Psychologists Press.

Form E of the SDS is a shortened form of the original version and was designed for students as young as the fourth grade as well as for adults with limited reading skills.

Self-Observation Scales by A. Jackson Stenner and William G. Katzenmeyer; c1974; Kindergarten-Grade 6; National Testing Service, Inc.

The instrument is designed to measure the way children perceive themselves and their relationships to peers, teacher, and school. The scales are available at two levels. Primary (K-3) assesses five dimensions of children's self-concept: Self-Acceptance, Social Maturity, School Affiliation, Self-Security, and Achievement Motivation. Intermediate (4-6) covers these five dimensions in addition to Social Confidence, Teacher Affiliation, and Peer Affiliation. Spanish versions are available at each level. Split-half reliability coefficients, corrected by Spearman-Brown Prophecy Formula, ranged from .75 to .81 for the subtests of the Primary Level and from .73 to .85 for those of the Intermediate Level. Validity is discussed in Technical Bulletins No. 1 and 2.

Self-Perception Inventory by William T. Martin; c1969; Grades 9-16; Psychologists and Educators Press.

This self-report inventory encompasses a number of important dimensions of personality: Uncommon Response, Time Score, Consistency, Self-Actualization, Supervision, Rigidity-Dogmatism, Authoritarianism, Anxiety, Depression, Paranoia, General Adjustment, and General Maladjustment. Reliability and validity data are presented in the manual.

Sequential Tests of Educational Progress: Original Series; c1956-65; Grades 4-14; Addison-Wesley Publishing Company, Inc.

This battery provides a series of achievement tests in the major

academic areas. Six tests are available at each of four levels covering grades 4 through 14. Kuder-Richardson Formula 20 was used to estimate reliabilities of all tests at all levels. The coefficients ranged from .74 to .95.

Listening

Tests ability to listen to and understand such materials as directions and simple explanations, exposition, narration, and poetry.

Mathematics

Measures mastery of number and operation, symbolism, measurement and geometry, function and relation, deduction and inference, and probability and statistics.

Reading

Measures ability to read and understand a variety of materials, requiring identification of details, analysis and interpretation, and criticism.

Science

Evaluates ability to identify and define scientific problems, test hypotheses, and interpret data. It includes questions in biology, chemistry, physics, meteorology, astronomy, and geology, with emphasis on practical applications.

Social Studies

Tests ability to analyze relationships among basic facts, trends, and concepts. It includes items in history, geography, economics, government, and sociology.

Writing

Evaluates ability to correct errors in writing organization, appropriateness, effectiveness, and grammar.

Sequential Tests of Educational Progress: Series II; c1956-72; Grades 4-14; Addison-Wesley Publishing Company, Inc.

This battery of achievement tests is designed to evaluate student progress toward fulfilling the broad, general goals of education. The six tests are available at four levels covering grades 4 through

14. Kuder-Richardson 20 reliabilities were computed for the tests at all levels. The coefficients were in the .80's and .90's:

English Expression

Assesses the ability to evaluate the correctness and effectiveness of sentences. The tests consist of two parts. Correctness of Expression stresses the application of the student's English proficiency rather than ability to use formal terminology. Effectiveness of Expression requires the student to demonstrate his sensitivity to language by selecting the answer that is grammatically correct, best in sentence structure and word order, and most precise and appropriate in idiom and diction.

Mathematics Basic Concepts

Measures elementary mathematical concepts, abilities, and skills. The concepts involve number and operation; measurement and geometry; relations, functions, and graphs; proof; probability and statistics; mathematical sentences; sets and mathematical systems; and application.

Mathematics Computation

Measures a wide variety of computational skills, including fundamental operations with integers, fractions, decimals, and percent; estimation; evaluation of formulas; solution of simple inequalities; and manipulations with exponents.

Mechanics of Writing

Evaluates the mastery of the fundamental composition skills of spelling, capitalization, and punctuation by asking students to identify misspelled words and to detect errors in capitalization and punctuation in the context of given sentences.

Reading

Measures the ability to read and understand a variety of materials. The sentences and passages that form the stimuli include stories and poems as well as selections from the literature of the sciences, social studies, and humanities. Part I contains sentence comprehension items of two basic types: straightforward comprehension and inference. Part II contains six passages of varying lengths. Several questions are asked about each passage.

Science

Measures knowledge and understanding of the fundamental concepts and processes of science, the application of this knowledge, and the mastery of science skills. Major emphasis is placed on biology; physics and chemistry receive moderate emphasis; and the earth sciences, including astronomy, geology, and meteorology, receive the least emphasis.

Social Studies

Measures development of the skills and knowledge requisite for effective citizenship. The skills include the ability to read, organize, interpret, and evaluate information presented in tabular, pictorial, or prose form. The knowledge includes the nature of social change; the interdependence of individuals, communities, and societies; the way in which society directs and regulates the behavior of its members; the nature of a democratic society; the effects of the environment on man's way of life and man's increasing control over the forces of nature; and man's economic needs.

Similes Test by Charles E. Schaefer; c1971; Grades 4-16 and Adults; Research Psychologists Press.

This test employs the ability to produce similes as an important factor underlying creative achievement. Each subject completes a total of ten simile stems by providing three different responses for each stem. The items pertain to various sensations (sight, sound, touch) and feelings (sadness, loneliness). The test is scored for degree of originality. Extensive scoring instructions are provided. Odd-even reliability estimates adjusted by Spearman-Brown formula, range from .82 to .89 for different samples. Two validity studies are described in the manual.

Situational Preference Inventory by Carl N. Edwards; 1968-73; Grades 9-16 and Adults; Carl N. Edwards.

The self-administering inventory was developed to assess individual styles of social interaction. It consists of 22 sets of statements. Each set contains three statements, each representing a different style of interaction: cooperational, instrumental, and analytic. The respondents are asked to indicate which of the three statements they agree with most and which they agree with least. Test-retest reliabilities for the three scales are: Cooperational .74, Instrumental .82, and Analytic .80. Validity data are presented.

Smith-Sturgeon Conditional Reasoning Test by Joanne Sturgeon and Wilson Smith; 1969; Ages 6-9; Cornell Critical Thinking Project.

An assessment of children's logical abilities, the instrument is concerned with four main principles: Inversion, Conversion, Contraposition, and Transitivity. Subscores can be obtained for each of the four principles, along with the fallacy principles and validity principles, the two forms of the questions (the suppositional and the factual), and for each of the two parts of the test (the house part and the chemicals part). Kuder-Richardson coefficients calculated ranged from .73 to .91 for all grades over all principles. Content and construct validity information is available.

Social and Prevocational Information Battery by Andrew Halpern, Paul Raffeld, Larry K. Irwin, and Robert Link; c1975; Grades 7-12; CTB/McGraw-Hill.

This battery is designed to assess knowledge of skills and competencies regarded as important for the community adjustment of Educable Mentally Retarded students. It consists of nine tests within five long range educational goals: Employability (Job Search Skills, Job Related Behavior), Economic Self-Sufficiency (Banking, Budgeting, Purchasing Habits), Family Living (Home Management, Physical Health Care), Personal Habits (Hygiene, Grooming), and Communication (Functional Signs). Kuder-Richardson formula 20 was used to compute reliability coefficients for samples of junior high school and senior high school students. They were .94 and .93 respectively. Test-retest reliabilities for predictive and concurrent validity are presented in the technical report.

Stanford Achievement Test: 1973 Edition by Richard Madden, Eric F. Gardner, Herbert C. Rudman, Bjorn Karlsen, and Jack C. Merwin; c1972-73; Grades 1.5-9.5; The Psychological Corporation.

This series of comprehensive achievement tests was developed to provide a measurement and assessment of learning at different levels of the educational process. The test is organized in six levels for various grades and covers various areas of achievement as follows:

Primary Level I is designed for use from the middle of grade 1 to the middle of grade 2 but may be used at the end of grade 2 for classes with a wide range of ability. It includes six single tests and one optional test: Vocabulary, Reading: Word Reading and Reading Comprehension, Word Study Skills, Mathematics Concepts, Mathematics Computation and Applications, Listening Comprehension, and optional Spelling test. Reliability coefficients based on split-half, odd-even scores corrected by the Spearman-Brown Formula ranged from .81 to .95. The same range was determined by the Kuder-Richardson Formula 20. Validity is discussed in the manual.

Primary Level II, grades 2-3, includes 11 tests: Vocabulary, Reading: Word Reading and Reading Comprehension, Word Study Skills, Mathematics Concepts, Mathematics Computation, Mathematics Applications, Spelling, Language, Social Science, Science, and Listening Comprehension. Split-half reliability coefficients based on odd-even scores corrected by the Spearman-Brown formula ranged from .69 to .96. K-R 20 coefficients ranged from .67 to .95. Validity information is presented in the manual.

Primary Level III, grades 3-4, includes 11 tests: Vocabulary, Reading Comprehension, Word Study Skills, Mathematics Concepts, Mathematics Computation, Mathematics Applications, Spelling, Language, Social Science, Science, and Listening Comprehension. Split-half estimates based on odd-even scores corrected by the Spearman-Brown Formula ranged from .86 to .96. K-R 20 reliability coefficients ranged from .86 to .96 also. Validity information is presented in the manual.

Intermediate Levels I and II, grades 4-5 and grades 5-6 respectively, include 10 tests at each level: Vocabulary, Reading Comprehension, Word Study Skills, Mathematics Concepts Computation, Mathematics Applications, Spelling, Language, Social Science, Science, and Listening Comprehension. For Level I, split-half, odd-even scores corrected by the Spearman-Brown formula ranged from .87 to .96; K-R 20 coefficients ranged from .89 to .95. For Level II, corrected scores by the Spearman-Brown Formula ranged from .85 to .95; K-R 20 estimates ranged from .85 to .95. Validity information is presented in the manuals.

Advanced Battery, grades 7-9, includes the same subtests as Intermediate Levels I and II except for Listening Comprehension. In the junior high school grades it is no longer meaningful to test listening as a separate ability. Reliability coefficients based on split-half, odd-even scores corrected by the Spearman-Brown Formula ranged from .87 to .95. The Kuder-Richardson Formula 20 yielded reliability coefficients ranged from .86 to .94. Validity information is discussed in the manual.

Stanford Diagnostic Arithmetic Test by Leslie S. Beatty; Richard Madden, and Eric F. Gardner; c1966-68; Grades 2.5-8.5; The Psychological Corporation.

Designed to identify needed areas of instruction in arithmetic, the test is intended for use in the early part of some instructional sequence. Tests are provided at two levels. Level I is intended for use with grades 2.5 to 4.5 and covers Concepts of Numbers and Numerals, Computation, and Number Facts. Level II is appropriate for grades 4.5 to 8.5. It covers Concepts of Numbers and Numerals, Computation with Whole Numbers, Common Fractions, Decimal Fractions and Per Cent, and Number Facts. Corrected split-half reliability coefficients ranged from .77 to .97 for Level I and from .57 to .96 for Level II. Validity data are presented in the manual.

Stanford Diagnostic Mathematics Test: 1976 Edition by Leslie S. Beatty, Richard Madden, Eric F. Gardner, and Bjorn Karlsen; c1976; Grades 1.5-13; The Psychological Corporation.

The test measures competence in the basic concepts and skills that are important in daily affairs and prerequisite to the continued study of mathematics. Designed to diagnose pupils' specific strengths and weaknesses, it provides tests at four levels: Red (1.5-4.5), Green (3.5-6.5), Brown (5.5-8.5), and Blue (7.5-13). The tests assess three interrelated areas of mathematical competence: Number System and Numeration, Computation, and Application. Kuder-Richardson Formula 20 was used to determine reliability of total test scores. The coefficients ranged from .94 to .97 for the various levels of the test. The content and criterion-related validity of the tests is discussed in the Manual for Administering and Interpreting.

Stanford Diagnostic Reading Test: 1976 Edition by Bjorn Karlsen, Richard Madden, and Eric F. Gardner; c1976; Grades 1.5-9.5; The Psychological Corporation.

The test is designed to measure those major components of the reading process considered to be prerequisite to successful reading behaviors. At the primary and elementary grades, reading can be conceptualized

as consisting of three major subskills: decoding, vocabulary, and comprehension. At the intermediate and secondary grades, reading rate also becomes important. Tests are available at three levels: Red (1.5-3.5), Green (2.5-5.5), and Brown (4.5-9.5). Kuder-Richardson Formula 20 was used to determine the internal consistency reliability of the subtest scores. Coefficients ranged from .82 to .97. Content and criterion-related validity of the test is discussed in the Manual for Administering and Interpreting.

Stanford Early School Achievement Test by Richard Madden and Eric F. Gardner; c1969-71; Kindergarten-Grade 1; The Psychological Corporation.

Designed to measure children's cognitive abilities, the test is available at two levels. Level I aids in the assessment of the background of pupils upon their entrance to school, and it helps to establish the baseline where instructional experiences should begin. Level II provides for more precise measurement of cognitive growth. The tests consist of six parts: The Environment, Mathematics, Letters and Sounds, Aural Comprehension, Word Reading, and Sentence Reading. Kuder-Richardson Formula 20 reliability coefficients were computed for each part of the test and range from .61 to .97. Validity data are presented in the Directions for Administering.

Stanford Test of Academic Skills (TASK) by Eric F. Gardner, Robert Callis, Jack C. Merwin, and Richard Madden; c1972-73; The Psychological Corporation.

This test at each of two levels, contains three tests which measure mastery of the basic skills: reading, English, and mathematics. Level I is designed for use in grades 8, 9, and 10; Level II for grades 11 and 12 and with junior/community college freshmen. The Reading Test measures reading comprehension and vocabulary. The English Test measures the student's knowledge and effective use of the English language. It covers skills such as dictionary use, reference sources, and the nature and structure of language; errors in capitalization, grammar, or punctuation; spelling; English expression; and logical sequence. The Mathematics Test measures general competence, emphasizing arithmetical and numerical concepts, computation, and applications. Split-half estimates based on odd-even items corrected by Spearman-Brown formula ranged from .92 to .96. Kuder-Richardson Formula 20 coefficients ranged from .92 to .95. Validity is discussed in the manual.

Strong-Campbell Interest Inventory by David P. Campbell and E.K. Strong; c1933-74; Ages 16 and Above; Stanford University Press.

This interest inventory is used in making educational and occupational choices. The interest blank is divided into seven parts: Occupations, School Subjects, Activities, Amusements, Types of People, Preference Between Two Activities, and Your Characteristics. The profile contains a variety of scores based on six General Occupational Themes (Realistic, Conventional, Enterprising, Investigative, Artistic, and Social); 23 Basic Interest Scales; and 124 Occupational Scales. Test-retest reliability over a 30 day period ranges from .83 to .93. Content, concurrent, and predictive validity are discussed in the manual.

Survey of Interpersonal Values by Leonard V. Gordon; c1960-63; Grades 9-16 and Adults; Science Research Associates, Inc.

The survey is designed to measure certain critical values involving the individual's relationships to other people or their relationships to him. These values are important in the individual's personal, social, marital, and occupational adjustment. The six values measured are Support, Conformity, Recognition, Independence, Benevolence, and Leadership. The instrument is self-administering. Test-retest reliability coefficients for the six scales range from .78 to .88. Reliabilities estimated by the Kuder-Richardson formula ranged from .71 to .86. Descriptive validity studies are discussed in the manual.

Survey of Personal Values by Leonard V. Gordon; c1964-67; Grades 9-16 and Adults; Science Research Associates, Inc.

The instrument is designed to measure certain critical values that help determine the manner in which an individual copes with the problems of everyday living. The six values are: Practical Mindedness, Achievement, Variety, Decisiveness, Orderliness, and Goal Orientation. The survey is self-administering and employs a forced-choice format. Test-retest reliabilities for the six scales range from .74 to .92. Kuder-Richardson reliability coefficients range from .72 to .92. Validity of the instrument is discussed in the manual.

Survey of School Attitudes by Thomas P. Hogan; c1973-75; Grades 1-8; The Psychological Corporation.

The survey is designed to measure student reactions to four major areas of school curriculum: Reading and Other Language Arts, Mathematics, Science, and Social Studies. Children indicate whether they like, dislike, or are neutral toward different activities in these areas. Two levels of the survey are available. The Primary Level is designed for grades 1 through 3 and for grade 4 students who have reading problems. At this level, the teacher reads the item "stems" to the students so that reading by students is not required. At the Intermediate Level, for grades 4 through 8, the teacher gives preliminary instructions, but students read and respond to the items independently. Corrected split-half reliability estimates for the four scales range from .82 to .87 for Primary Level and from .83 to .90 for Intermediate Level. Coefficient Alpha coefficients range from .79 to .85 for the Primary and from .84 to .92 for the Intermediate Level. Validity data are presented and discussed in the manual.

Survey of Study Habits and Attitudes High School Form by William F. Brown and Wayne H. Holtzman; c1953-67; Grades 7-12; The Psychological Corporation.

The survey was developed to measure study methods, motivation for studying, and certain attitudes toward scholastic activities which are important in the classroom. It yields seven scores: Delay Avoidance, Work Methods, Study Habits, Teacher Approval, Education Acceptance, Study Attitudes, and Study Orientation. A Spanish version of the survey is available. Reliability coefficients obtained for the subscales using the Kuder-Richardson Formula 8 ranged from .87 to .89. Validity of the survey is discussed extensively in the manual.

Taylor-Helmstadter Pair Comparison Scale of Aesthetic Judgment by Anne P. Taylor and G.C. Helmstadter; c1973; Ages 4 Years and Above; University Testing Service.

The scale consists of 38 pairs of color slides depicting paintings, sculpture, and common household items. One slide of each pair was

judged higher than the other on a scale of aesthetic quality by ten art experts. The examinee is asked to choose the one of each pair of slides that he likes better. The score is the number of choices that correspond to the preference of the art experts. Studies of reliability and validity are in progress.

Tennessee Self-Concept Scale by William H. Fitts; c1964-70; Ages 12 and Over; Counselor Recordings and Tests.

This instrument is designed to measure self-concept as a means of studying and understanding human behavior. It is available in two forms: a Counseling Form and a Clinical and Research Form. It consists of 100 self descriptive statements which the subject uses to portray his own picture of himself. Subscores include: Self-Criticism, Net Conflict, Total Conflict, Positive, Identity, Self Satisfaction, Behavior, Physical Self, Moral-Ethical Self, Personal Self, Family Self, Social Self, and Variability. Test-retest reliability coefficients range from .70 to .92. Content validity has been examined and is discussed in the manual.

Test of Creative Potential by Ralph Hoepfner and Judith Hemenway; c1973; Ages 7-Adults; Monitor.

This short, comprehensive measure of general creative potential assesses the creativity factors of fluency, flexibility, originality, and elaboration. It consists of three subtests. Writing Words can be scored for associational fluency or for spontaneous flexibility. Picture Decoration yields a score for degree of decorative elaboration. License Plate Words can be scored for word fluency or for originality. Inter-scorer reliability estimates range from .94 to .97 for different grade levels. Internal consistency reliability coefficients range from .65 to .79. Validity of the test is discussed briefly in the manual. Additional validity studies are being conducted.

Test of Elementary Economics; c1971; Grades 4-6; Joint Council on Economic Education.

This experimental instrument was developed to assess the effectiveness of an economic education curriculum. It evaluates students' knowledge, comprehension, and application of economic concepts and principles in the following areas: Household, Business, Government, Exchange, Technology, Market, and National Economy. Technical information is not available in the manual.

Test of Social Insight: Youth Edition by Russell N. Cassel; c1959-63; Grades 6-12; Martin M. Bruce, Publisher.

The test assesses the characteristic mode of reaction the individual uses in resolving personal social problems. Each of the 60 items is accompanied by five alternatives. Each alternative offers one of the following modes of resolution of the problem expressed in the item: withdrawal, passivity, cooperativeness, competitiveness, and aggressiveness. The social problems covered fall into four spheres of typical life activity: Home and Family Relations, Authority Figures and Social Agencies, Play and Avocational Interests, and Work and Vocational Interests. Split-half reliabilities for various samples range from .50 to .83. A test-retest reliability coefficient of .83 was obtained. Concurrent validity data are presented in the manual.

Test of Understanding in Personal Economics; c1971; Grades 9-12; Joint Council on Economic Education.

This test is designed to determine the effectiveness of courses intended to develop student competency in personal economics. The items emphasize the application of economic analysis to personal decisions and the interrelatedness of economic matters, both personal and social. Kuder-Richardson Formula 20 was used to estimate reliability coefficients for a variety of samples. The coefficient for the total group is .84. Validity of the test is not discussed.

Tests of Basic Experiences (TOBE) by Margaret H. Moss; c1970-72; Preschool-Grade 1; CTB/McGraw-Hill.

TOBE provides an indication of how well a child's experiences have prepared him for his introduction to many of the scholastic activities that he will encounter. TOBE is a pictorial, non-written test designed in two levels: Level K for preschool and kindergarten and Level L for Kindergarten and first grade. There is also a Spanish language version of the test. Subtests include: Mathematics, Language, Science, Social Studies, and General Concepts. Reliability was investigated by means of the Kuder-Richardson 20 formula. For Level K used with prekindergarten children, the coefficients ranged between .79 and .84 for the five subtests; with kindergarten children, the coefficients fell between .82 and .85. For Level L used with kindergarten children the coefficients ranged between .73 and .81; with first graders the coefficients fell between .78 and .82. Validity information is presented in the manual.

Thinking Creatively with Sounds and Words by E. Paul Torrance, Joe Khatena, and Burt F. Cunningham; c1973; Grades 3-12; Personnel Press.

This device is designed to assess creative thinking by measuring the originality of ideas stimulated by abstract sounds and spoken onomatopoeic words. Two independent tests are provided: Sounds and Images, and Onomatopoeia and Images. Both tests are available in alternate forms A and B for each of two levels, I (Grades 3-12) and II (Adults). A set of records is required to administer all of the tests. Split-half reliability coefficients, corrected by the Spearman-Brown prophecy formula range from .77 to .91 for Sounds and Images and from .76 to .95 for Onomatopoeia and Images. Content, construct, and criterion-related validity have been examined and data are presented in the manual.

Thomas Self-Concept Values Test by Walter L. Thomas; c1967-72; Ages 3-9; W. Clement and Jessie V. Stone Foundation.

The test was developed to provide a technique for assessing the personal self-concept values of young children. It allows the individual to reflect and report on himself as "subject" and as "object" in light of

commonly conceived value dimension--self-value indices. The 14 self-value factors included in this instrument are: Happiness, Cleanliness, Sociability, Sharing, Ability, Male Acceptance, Fear (of things), Fear (of people), Strength, Size, Health, Attractiveness, Material, and Independence. The child must indicate how he perceives himself and how he thinks significant others (mother, teacher, peers) perceive him in terms of these factors. Test-retest coefficient of reliability is .78 for the total score. Construct and concurrent validity is discussed.

Thurstone Interest Schedule by L.L. Thurstone; c1947; Grades 9-16 and Adults; The Psychological Corporation.

With this checklist, a person can systematically clarify his understanding of his vocational interests. It is designed as a counseling instrument to be used in situations in which the client-counselor relationship is such that straightforward and honest expression of choices can be expected. The 10 scores in the profile represent relative interest in 10 vocational fields as follows: Physical Science, Biological Science, Computational, Business, Executive, Persuasive, Linguistic, Humanitarian, Artistic, and Musical. Split-half reliability coefficients computed ranged from .90 to .96. Validity information is presented in the manual.

Torrance Tests of Creative Thinking by E. Paul Torrance; c1966; Kindergarten and Above; Personnel Press.

Two tests are provided to assess four characteristics that constitute creative thinking ability: Fluency, Flexibility, Originality, and Elaboration. The Verbal Test uses seven word-based exercises to measure these traits: Asking, Guessing Causes, Guessing Consequences, Product Improvement, Unusual Uses of Cardboard Boxes (or Tin Cans), Unusual Questions, and Just Suppose. The Figural Test uses three picture-based activities for the same purpose: Picture Construction, Picture Completion, and Lines (or Circles). A number of reliability studies have been conducted and the data are reported in the Norms-Technical Manual. Content and construct validity studies are also discussed.

Watson-Glaser Critical Thinking Appraisal by Goodwin Watson and Edward M. Glaser; c1951-64; Grades 9-16 and Adults; The Psychological Corporation.

This instrument consists of a series of test exercises that require the application of some of the important abilities involved in critical thinking. The exercises include problems, statements, arguments, and interpretation of data similar to those that an individual in a democracy encounters in his daily life. The instrument consists of five subtests designed to measure different, though interdependent, aspects of critical thinking. The five subtests are as follows: Inference, Recognition of Assumptions, Deduction, Interpretation, and Evaluation of Arguments. Split-half, odd-even reliability coefficients corrected by the Spearman-Brown formula range from .53 to .74 for Form Ym and .40 to .55 for Form Zm. Content, construct, and predictive validity information is presented in the manual.

Wide Range Interest and Opinion Test (WRIOT) by Joseph F. Jastak and Sarah R. Jastak; c1972; Grades K-16 and Adults; Guidance Associates of Delaware, Inc.

This inventory of work interests is designed to cover as many areas and levels of human activity as possible in pictures. The activities portrayed are representative of those listed in the Dictionary of Occupational Titles. They include a wide range of work from unskilled labor through technical occupations to professional and managerial positions. Vocational attitudes are also explored. The instrument is most useful in career planning, personnel work, guidance, and vocational rehabilitation. Cureton's formula for split-half reliability yielded coefficients ranging from .88 to .95 for males and females. The WRIOT must be validated daily by those who use it.

Woodcock Reading Mastery Tests by Richard W. Woodcock; c1973; Kindergarten-Grade 12; American Guidance Service, Inc.

This battery of criterion-referenced tests assesses reading skills in five areas: Letter Identification, Word Identification, Word Attack, Word Comprehension, and Passage Comprehension. While the test is designed to provide for criterion-referenced interpretation, raw scores

can be converted to traditional normative scores. In addition, separate norms are available for boys and girls as are SES adjusted norms based on communities having SES characteristics similar to the local community. Split-half reliabilities for total test score range from .97 to .99. Test-retest alternate form reliabilities range from .83 to .97. Evidence related to validity is drawn from four sources: content validity, a multitrait-multimethod matrix analysis, further intercorrelation data, and a predictive study using the Mastery Scale. All are discussed in the manual.

Work Values Inventory by Donald E. Super; c1968-70; Grades 7-16 and Adults; Houghton Mifflin Company.

The inventory was developed to assess the goals that underlie the motivation to work; that is, the values considered important in determining an individual's satisfaction and success in his vocation. The fifteen values measured by the inventory are: Intellectual Stimulation, Job Achievement, Way of Life, Economic Returns, Altruism, Creativity, Relationships with Associates, Job Security, Prestige, Management of Others, Variety, Aesthetics, Independence, Supervisory Relations, and Physical Surroundings. The inventory was normed in 1968 on a national sample of 10,083 students in grades 7 through 12. Test-retest reliability is reported for each of the scales. The coefficients range from .74 to .88 with a median of .83. The construct, content, concurrent, and predictive validity is examined in the manual.

Written School Attitude Test by Earl McCallon; c1973; Grades 4-6; Learning Concepts.

The test is intended to reflect a student's perception of his school environment. The 46 items are designed to elicit a response from the student that will be representative of his attitude toward his school environment and educational experiences. They deal with three dimensions: Interpersonal Relations, Student-Instruction Interaction, and General School Factor. Both Spanish and English forms of the test are available. Three test-retest reliability studies were conducted and the coefficients were .82, .75, and .78 for 10, 21, and 14 days, respectively. The test appears to have content validity.

Youth Research Survey by Merton P. Strommen and Ram K. Gupta; c1969-71; Grades 9-13; Youth Research Center.

This instrument is designed for use in the individual counseling of youth, in determining the needs of youth groups, and evaluating the impact of what is being done in a particular school or congregation. The 24 scales are classified into three major categories: Concerns, Beliefs-Values, and Perception. The Concern Scales measure: Family Unity, Parental Understanding, Family Pressures, Life Partner, Lack of Self-Confidence, Academic Problems, Personal Faults, Classroom Relationships, National Issues, and God Relationship. The Belief-Value Scales measure: Interest in Help, Maturity of Values, Orientation for Change, Moral Responsibility, Meaningful Life, Religious Participation, Social Action, Self-Regard, Human Relations, God Awareness, and Biblical Concepts. The Perception Scales measure Youth Group Vitality, Adult Caring, and Family Social Concerns. There is also a Validity Scale which measures Frankness. The Concern Scale reliability estimates range from .56 to .89. The Belief-Values Scale yields reliabilities ranging from .59 to .93. The Perception Scale yields an average reliability coefficient of .77. Validity information is discussed in the manual. A technical report is available from the Youth Research Center upon request.

APPENDIX A-4 - Publisher Addresses

- | | |
|--|---|
| (1) Addison-Wesley Publishing Company
South Street
Reading, Massachusetts 01867 | (13) Counselor Recordings and Tests
Vanderbilt University
Box 6184, Acklen Station
Nashville, Tennessee 37212 |
| (2) The American College Testing Program
P.O. Box 168
Iowa City, Iowa 52240 | (14) Critical Thinking Project
371 Education Building
University of Illinois
Urbana, Illinois 61801 |
| (3) American Guidance Service, Inc.
Publishers' Building
Circle Pines, Minnesota 55014 | (15) Dissemination Center for Bilingual
Bicultural Education
6504 Tracor Lane
Austin, Texas 78721 |
| (4) Arden Press
8331 Alvarado Drive
Huntington Beach, California 92646 | (16) Educational and Industrial Testing
Service
P.O. Box 7234
San Diego, California 92107 |
| (5) Dr. Marshall N. Arlin
Department of Educational Psychology
University of British Columbia
Vancouver, B.C., Canada V6T 1W5 | (17) Educational Improvement Center
Halko Drive
Cedar Knolls, New Jersey 07927 |
| (6) Association Press
291 Broadway
New York, New York 10007 | (18) Educational Skills Development, Inc.
179 Maxwell Street
Lexington, Kentucky 40508 |
| (7) Bureau for Educational Research
and Service
The University of Iowa
C-6, East Hall
Iowa City, Iowa 52240 | (19) Dr. Carl N. Edwards
P.O. Box 86
Village Station
Medway, Massachusetts 02053 |
| (8) Bureau of Educational Measurements
Kansas State Teachers College
Emporia, Kansas 66801 | (20) Family Life Publications, Inc.
219 Henderson Street
P.O. Box 427
Saluda, North Carolina 28773 |
| (9) CTB/McGraw-Hill
Del Monte Research Park
Monterey, California 93940 | (21) B. M. Farley
53 Edgewood Avenue
Little Silver, New Jersey 07739 |
| (10) Campus Publishers
Box 1005
Mankato, Minnesota 56001 | (22) Fearon Publishers
6 Davis Drive
Belmont, California 94002 |
| (11) Center for Psychological Service
Suite 419, Columbia Medical Building
1835 Eye Street, N.W.
Washington, D.C. 20006 | (23) Florida Educational Research and
Development Council
College of Education
University of Florida
Gainesville, Florida 32601 |
| (12) Consulting Psychologists Press, Inc.
777 College Avenue
Palo Alto, California 94306 | |

- (24) Follett Publishing Company
P.O. Box 5705
1010 West Washington Boulevard
Chicago, Illinois 60607
- (25) Ginn and Company
Education Center
P.O. Box 2649
Columbus, Ohio 43216
- (26) Guidance Associates of Delaware, Inc.
1526 Gilpin Avenue
Wilmington, Delaware 19806
- (27) Houghton Mifflin Company
110 Tremont Street
Boston, Massachusetts 02107
- (28) Institute for Personality and Ability
Testing
1602 Coronado Drive
Champaign, Illinois 61802
- (29) The Instructional Objectives Exchange
Box 24095
Los Angeles, California 90024
- (30) Joint Council on Economic Education
1212 Avenue of the Americas
New York, New York 10036
- (31) Dr. Joe Khatena
Department of Educational Foundations
Marshall University
Huntington, West Virginia 25701
- (32) Alma Jordan Knauber
9871 Lorelei Drive
Cincinnati, Ohio 45231
- (33) Learning Concepts, Inc.
2501 N. Lamar
Austin, Texas 78705
- (34) Karen A. Maitland
Department of Psychology
University of Florida
Gainesville, Florida 32601
- (35) Martin M. Bruce, Publishers
340 Oxford Road
New Rochelle, New York 10804
- (36) Mill Music, Inc.
1619 Broadway
New York, New York 10019
- (37) Monitor
P.O. Box 2337
Hollywood, California 90028
- (38) N.F.E.R. Publishing Company, Ltd.
"the Mere", Upton Park
Slough, Bucks
Great Britain
- (39) National Testing Service, Inc.
2526 Erwin Road
Durham, North Carolina 27705
- (40) Dr. Clarence Nelson
Office of Evaluation Services
Michigan State University
202 S. Kedzie Hall
East Lansing, Michigan 48824
- (41) Personnel Press
Education Center
P.O. Box 2649
Columbus, Ohio 43216
- (42) Person-O-Metrics
205-04 Williamsburg Road
Dearborn Heights, Michigan 48127
- (43) Peters and Associates
328 Huckleberry Hill Road
Avon, Connecticut 06001
- (44) Phi Delta Kappa
Center for the Dissemination of
Innovative Programs
8th and Union, Box 789
Bloomington, Indiana 47401
- (45) The Psychological Corporation
757 Third Avenue
New York, New York 10017
- (46) Psychological Publications, Inc.
5300 Hollywood Boulevard
Los Angeles, California 90027
- (47) Psychologists and Educators, Inc.
Suite 212
211 West State
Jacksonville, Illinois 62650
- (48) Psychometric Affiliates
Box 3167
Munster, Indiana 46321

- (49) Research Concepts
A Division of Test Maker, Inc.
1368 East Airport Road
Muskegon, Michigan 49444
- (50) Research Psychologists Press, Inc.
36 St. John Street
Goshen, New York 10924
- (51) Dr. James R. Rest
University of Minnesota
330 Burton Hall
Minneapolis, Minnesota 55455
- (52) San Felipe Press
Box 2085
Austin, Texas 78767
- (53) San Diego County
Department of Education
6401 Linda Vista Road
San Diego, California 92111
- (54) Scholastic Testing Service, Inc.
480 Meyer Road
Bensenville, Illinois 60106
- (55) Science Research Associates, Inc.
259 East Erie Street
Chicago, Illinois 60611
- (56) Self Esteem Institute
1736 Stockton Street
San Francisco, California 94133
- (57) Stanford University Press
Stanford, California 94305
- (58) Dr. Donald E. Super
Department of Counseling and
Personnel Psychology
Teachers College
Columbia University
New York, New York 10027
- (59) Teachers College Press
Teachers College
Columbia University
525 West 120th Street
New York, New York 10027
- (60) Tests in Microfiche (TIM)
Test Collection
Educational Testing Service
Princeton, New Jersey 08540
- (61) University Testing Service
Payne Hall, B302
Arizona State University
Tempe, Arizona 85281
- (62) W. Clement and Jesse V. Stone
Foundation
Achievement Motivation Program
111 East Wacker Drive
Suite 510
Chicago, Illinois 60601
- (63) Western Psychological Services
12031 Wilshire Boulevard
Los Angeles, California 90025
- (64) Westinghouse Learning Corporation
P.O. Box 30
Iowa City, Iowa 52240
- (65) Wisconsin Research and Development
Center for Cognitive Learning
1025 West Johnson Street
Madison, Wisconsin 53706
- (66) Youth Research Center
122 West Franklin Avenue
Minneapolis, Minnesota 55404

APPENDIX B-1 - Selecting a Sample and Interpreting
Assessment Results from the Sample

Purpose. The purpose of this section is to present some rules of thumb for interpreting certain numbers, such as averages. The section will include a discussion on how to tell if differences between averages are likely to represent real or "significant" differences.

Sampling. For many district-level evaluations it is not necessary to test every student. To save time and money, it is reasonable to test only a portion or sample of the students, and then generalize from results on the sample to the results that would have been obtained if everybody had been tested. The process works to the extent that the sample selected represents the population being sampled.

One of the best ways to get representation is to select a "random" sample. In a random sample, students are picked purely by chance so that each student has an equal chance to be chosen. For example, if you wanted a random sample of the second-grade students in a district you could write the name of each student on a slip of paper, throw all the slips into a box, and pick out names with your eyes shut. There are, of course, many ways to select random and other types of samples. They are explained in references cited in Appendix C-2.

Sampling Error. Whenever you use the results of a sample to talk about a whole group you are in danger of being in error. It is very unlikely that the results from a sample will be exactly the same as the results obtained for the whole group. In fact, each sample you pick from a group

is likely to be a bit different. There will be what is called "sampling error."

A way around the problem is to figure out how much error is likely to occur in the sample results. You can then say, "I don't know the exact value for the whole group, but based on the sample it's likely to be between these two values."

How to Compute Sampling Error. To compute sampling error you need only two pieces of information: (1) the size of the sample and (2) the "standard deviation." The standard deviation is a number that tells how varied or spread-out or different from each other a group of numbers are. There are many calculators selling for less than \$100 that will compute a standard deviation at the touch of a button. If you have to calculate one by hand read the last page of this section.

To find the sampling error, just divide the standard deviation of the sample by the square root of the sample size. This gives what is called the standard error. Expressed as an equation:

$$\text{STANDARD ERROR} = \frac{\text{S.D}}{\sqrt{N}}$$

For example, as part of a school's health program you may want to know the average number of dental cavities to be found in high school students. Let's say there are about 3,000 high school students in the district, but you can afford to have only 150 examined by a dentist. You select a random sample of 150 students and find that the average number of cavities is 3.44. A few minutes with a calculator will tell you that the standard deviation is 1.87. What can you say about the high school students on the basis of the sample?

First you compute the standard error:

$$\text{Standard Error} = \frac{S.D}{\sqrt{N}} = \frac{1.87}{\sqrt{150}} = \frac{1.87}{12.25} = .15$$

Then you can make statements based on the following table:

WHOLE GROUP AVERAGE WILL BE IN THE INTERVAL FORMED BY THE SAMPLE AVERAGE PLUS AND MINUS:		
1	standard error	68% of the time
1.96	standard error	95% of the time
2	standard errors	96% of the time
2.58	standard errors	99% of the time

For our example if you want to be 68 percent sure of being right you could say that the average number of cavities per high school student is between 3.29 and 3.59 ($3.44 - .15$ to $3.44 + .15$). If you want to be 95 percent sure of being right you could say that the average is between 3.146 and 3.734 [$3.44 - 1.96 (.15)$ to $3.44 + 1.96 (.15)$]. If you want to be 99 percent sure you could say the average is between 3.053 and 3.827 [$3.44 - 2.58 (.15)$ to $3.44 + 2.58 (.15)$]. For those who care to talk statistically, using a band of plus and minus 1.96 standard errors puts you at the ".05 level of significance." Using plus or minus 2.58 standard errors puts you at the ".01 level of significance."

Comparing numbers. Let's say we start a dental health program in the high school. After a year we select another random sample of 150 students and count their cavities. We get an average of 3.01 and a standard deviation of 1.42. Can we say that there is a real difference from one year to

the next? Or would a difference that size be found just because of sampling error?

We know what the sampling error was for the first year. We calculated it as $\frac{SD}{\sqrt{N}} = \frac{1.87}{\sqrt{150}} = \frac{1.87}{12.25} = .15$. But the second year average also has a standard error that we have to calculate. $\frac{SD}{\sqrt{N}} = \frac{1.42}{\sqrt{150}} = \frac{1.42}{12.25} = .116$. Note that the error of comparison will be larger than the error of either sample.

Combining Errors. It is tempting to just add up the two errors, but that won't work. This is what has to be done:

- (1) square each standard error
- (2) add up the squared errors
- (3) take the square root of the sum.

For our example:

square error 1	$.15 \times .15 = .0225$
square error 2	$.116 \times .116 = .0135$
add them up	$.0225 + .0135 = .036$
take the square root	$\sqrt{.036} = .19$

Once you get the combined error it is easy to determine if the difference between means is likely to be real or significant. Just follow these steps:

- (1) subtract one mean from the other
- (2) divide the difference by the combined standard error
- (3) If the ratio is 1.96 or larger it is significant at the .05 level or beyond. That means that you are about 95 percent sure that the difference is real. If the ratio is 2.58 or larger, the difference is significant at the .01 level or

beyond. You can be 99 percent sure that the difference is a real one.

If the ratio turns out to be less than 1.96, most statisticians would be unwilling to call the difference "significant."

For our example:

- (1) subtract one mean from the other $3.44 - 3.01 = .43$
- (2) divide difference by combined standard error. $.43/.19 = 2.26$
- (3) is ratio larger than 1.96? yes
- (4) is ratio larger than 2.58? no
- (5) The difference is significant at the .05 level. You are safe in saying that there is a real difference in the number of cavities.

The formula for calculating a standard deviation looks hard, but the individual steps are easy.

$$\text{STANDARD DEVIATION} = \sqrt{\frac{\sum (x - \bar{x})^2}{N - 1}}$$

- (1) write down each number (x)
- (2) add them up and divide by the number of numbers to get the average ($\bar{x}$)
- (3) subtract the average from each number to get a difference ($x - \bar{x}$)
- (4) square each difference [$(x - \bar{x})^2$]
- (5) add up the squared differences [$\sum (x - \bar{x})^2$]
- (6) subtract 1 from the number of numbers N-1
- (7) divide the added-up squared differences by the number of numbers minus 1 [$\frac{\sum (x - \bar{x})^2}{N - 1}$]
- (8) take the square root of the result $\sqrt{\frac{\sum (x - \bar{x})^2}{N - 1}}$

Worked Out Example of Standard Deviation

EACH NUMBER	THE AVERAGE	EACH NUMBER MINUS THE AVERAGE	SQUARE THE DIFFERENCES
x	$\bar{x}$	(x - $\bar{x}$)	(x - $\bar{x}$) ²
2	4	-2	4
6	4	2	4
3	4	-1	1
7	4	3	9
5	4	1	1
1	4	-3	9

28 ADD UP THE SQUARED
DIFFERENCES $\sum (x - \bar{x})^2$

Subtract 1 From Number of Numbers (N-1) = 6 - 1 = 5

Divide the Added Up Squared Differences by
the Number of Numbers Minus 1 $\frac{28}{5} = 5.6$

$$\frac{\sum (x - \bar{x})^2}{N - 1}$$

Take the Square Root of the Result $\sqrt{5.6} = 2.36$

$$\sqrt{\frac{\sum (x - \bar{x})^2}{N - 1}}$$

STANDARD DEVIATION = 2.36

For other methods of calculation see the statistics books cited in Appendix C-2.

APPENDIX B-2 - Some Short-Cut Statistics
for Internal Measures

The following material is information from Paul B. Diederich's Short-Cut Statistics for Teacher-Made Tests (Princeton, N.J.: Educational Testing Service, 1973).

If you have a measure for which there should be a right answer, you can determine how well it discriminates, whether the measure clearly separates those that are successful from those that are not by a short-cut item analysis. This is done by scoring the measure and then arranging the papers in descending order of total scores, counting down to the middle score. Suppose this score is 20, and five students received it. All papers above this score go into the high group; those below go into the low group. Separate the papers into two piles on this basis. Also, put the middle score papers at random into the two piles until the numbers of papers in each pile are equal. If you have an odd number of students, hold out one middle paper and do not count it in the item analysis. Now, distribute the papers to the students so that those with the "highs" are separated from those with the "lows." Call out the number of each item one by one, having students with that item right on the paper they are holding raise their hands. The four figures you will obtain are:

H = the number of highs who got the item right

L = the number of lows who got the item right

H+L = successful students, the total number who got the item right

H-L = discrimination indicator, the high-low difference or how many

more "highs" than "lows" got the item right

Research has shown that measurement is most precise when all items are about equally difficult for the group on which data is collected. This is when all items on the usual multiple-choice measure can be answered correctly by 60 percent of the students. However, this is often not possible. Therefore, most measures have a wide range of item-difficulties. For most measures other than criterion-referenced, to judge which items should be included, use the following guidelines:

- (1) Items that more than 90 percent got right should be questioned as too easy and items that fewer than 30 percent got right as too hard for inclusion.
- (2) Exclude items for which the high-low difference is less than 10 percent of the class. For example, in a class of 30 students, at least three more students in the top half than in the bottom half got the item right.

For a criterion-referenced measure, an item is judged good if it is passed by a high percentage of a post-instruction group if the instruction has been effective. That is, items which are difficult for students prior to instruction but easy after instruction are the ones that should be included in a criterion-referenced measure.

Besides needing to know the discrimination power of items in a measure, as described above, you should also be concerned about the extent to which the scores may be expected to vary by pure chance in the selection of items. For this we look at the standard error of scores. Use the following table to estimate the standard error of test scores:

NUMBER OF ITEMS	STANDARD ERROR	EXCEPTIONS: Regardless of the length of tests the standard error is:
24	2	0 when the score is zero or perfect;
21-47	3	1 when 1 or 2 points from 0 or from 100%;
48-89	4	2 when 3 to 7 points from 0 or from 100%;
90-109	5	3 when 8 to 15 points from 0 or from 100%.
110-129	6	
130-150	7	

The table may be interpreted as follows: In an objective test of 60 items, two scores out of 3 will lie within 4 raw-score points (one standard error) of the true score these students would attain if you continued testing repeated random samples from the universe of items testing the same ability, and 95 percent of the scores will be within 8 raw-score points (two standard errors) of true score. The relatively few scores of the extremes (3-7 percent correct and 93-100 percent correct responses) will have slightly smaller standard errors, as indicated under "Exceptions."

If we add to the standard error of scores our own bias in the selection of items, the mistakes we make in writing and scoring them, and external circumstances that may affect students' ability to respond, it is obvious that the difference between two independent measures of an ability that we refer to by a single objective or name may be quite large. This also reemphasizes the need for a broad data base.

Another concern should be with the reliability of the measure, the extent to which a measure consistently measures what it is said to measure. Reliability can be calculated by statistical methods, several of which are found in references listed in Appendix C-2. However, if you can also

administer the same measure twice to the same group of students (a test-retest procedure) and then examine the results to see if they are consistent, you will have a measure of reliability. To compare the results: find the percentage of students who stood in the top half of the group on both administrations and look up the correlation (r) corresponding to this percentage in the following table:

<u>%</u>	<u>r</u>	<u>%</u>	<u>r</u>	<u>%</u>	<u>r</u>	<u>%</u>	<u>r</u>	<u>%</u>	<u>r</u>
45	.95	37	.69	29	.25	21	-.25	13	-.69
44	.93	36	.65	28	.19	20	-.31	12	-.73
43	.91	35	.60	27	.13	19	-.37	11	-.77
42	.88	34	.55	26	.07	18	-.43	10	-.81
41	.85	33	.49	25	.00	17	-.49	9	-.85
40	.81	32	.43	24	-.07	16	-.55	8	-.88
39	.77	31	.37	23	-.13	15	-.60	7	-.91
38	.73	30	.31	22	-.19	14	-.65	6	-.93

Since 1 percent of the students can make a big difference in the correlation, it is important to use a standard method of counting the number of students who were in the top half on each administration of the measure. To find those in the top half, list all scores from highest to lowest and tally the number of students who made each score. After tallying all of the scores, count down the tallies to half the number of students in the group. This is the middle score. On the alphabetical listing of all students, circle the names of students with the middle score. Then identify those students whose score stood above the middle score by putting checks after their names. If half the students in the

class do not have checks after their names and you need two more students to make up half of the class, put a check after the names of the first two names circled on the list, the first two on the list with middle scores. Then count how many students have two checks after their names. Turn this number into a percent by dividing it by the total number of students. Find this percent in the table, and the decimal next to it will be the correlation between the two administrations of the measure. The higher the correlation, the more reliable is the measure. Generally, a correlation of over .85 should be obtained for a test-retest situation, although a reliability of above .60 is generally acceptable for this as well as other kinds of computed reliabilities.

APPENDIX B-3 - Evaluation of Program Effect in Student Achievement

This material is derived from some of the material prepared by G. Gasten Tallmadge and Christine T. Wood (Mountain View, California: RMC Corporation, 1976). This evaluation model is intended to determine how much more a group of students learns by participating in a program than by not participating in it. The relationship is expressed as follows:

$$\begin{array}{l} \text{Program Effect} \\ \text{(PE)} \end{array} = \begin{array}{c} \left[\begin{array}{c} \text{Observed} \\ \text{Post-program} \\ \text{Performance} \\ \text{(OPP)} \end{array} \right] - \begin{array}{c} \left[\begin{array}{c} \text{Expected} \\ \text{No-program} \\ \text{Performance} \\ \text{(ENP)} \end{array} \right] \end{array}$$

The observed post-program performance (OPP) is always indicated by the mean or median posttest score for the group. The expected no-program performance (ENP) is arrived at by one of three techniques:

- (1) Administer a test with national norms as a pre-program test and determine the percentile status of the performance at that time. Since it is assumed that the status of the group at post-program time would be the same, this pre-program test percentile status becomes the ENP. The same test or a parallel form is administered after the program has been implemented and the OPP is simply the percentile rank corresponding to the mean or median score.
- (2) Designate two comparable groups of students, one to participate in the program and the other to act as a control group by not participating. Administer a nationally normed measure

for pre- and posttesting of both groups. Between the pre- and posttesting, the educational experiences of both groups should be identical or very similar except that the control group does not participate in the program. The control group's posttest percentile or mean raw score is the ENP while the other group's posttest percentile or mean raw score is the OPP. The pretest scores of the groups are used only to verify the pretreatment comparability of the two groups.

- (3) This involves a regression-based evaluation design. Any measure, rating or composite measure that correlates highly with the posttest can be used as a pretest. The posttest is a normed test. Two groups are selected based on the pretest which has a cut-off score: those above the cut-off score are a control group (not in the program) while those below participate in the program. Post- and pretest regression lines are calculated for each group. The program participating group's regression line represents the OPP corresponding to various pretest scores. The control group's regression line, projected across the cut-off score, provides the ENP for the same pretest scores. The program effect (PE) is defined as the distance between the regression lines and is measured separately at two points: at the program participating group's mean pretest score and at the cut-off score. [Note: A substantial difference between the two measures may signal a spurious apparent gain due to test ceiling or floor effects.]

APPENDIX C-1 - Selected Sources of Test (External Measures) Reviews

The references annotated in this bibliography are sources of information on available measurement instruments. Some of the references cover commercially available tests, others include information on research measures, and a few list both research and commercial tests. In instances where the document is available from the ERIC Document Reproduction Service (EDRS), the appropriate ED number has been provided in the citation.¹

Beatty, Walcott H. Improving Educational Assessment and an Inventory of Measures of Affective Behavior. Washington, D.C.: Association for Supervision and Curriculum Development, National Educational Association, 1969.

The second section of the book, Inventory of Measures of Affective Behavior is an annotated list of measures in the following areas: attitudes, creativity, interaction, motivation, personality, readiness, self-concept, and miscellaneous.

Berger, Barbara. An Annotated Bibliography of Measurements for Young Children. New York: Center for Urban Education, 1969.

This is a selective listing of assessment instruments appropriate for preschool and kindergarten children. Included are research and commercially available measures of cognitive status, cognitive abilities, perceptual skills, reading readiness, characteristics of cognitive style, and personal-social development.

Blanton, William, Farr, Roger, and Tuinman, J. Jaap. Reading Tests for the Secondary Grades: A Review and Evaluation. Newark, Delaware: International Reading Association, 1972.

This book is intended primarily for classroom teachers and other personnel who are responsible for selecting reading achievement tests. Several of the most commonly used tests currently available for use at the high school level are reviewed and evaluated as to their content and statistical characteristics.

Bonjean, Charles M., Hill, Richard J., and McLemore, S. Dale. Sociological Measurement: An Inventory of Scales and Indices. San Francisco: Chandler Publishing Company, 1967.

This is an extensive bibliography of references for over 2,000 sociological scales and indices.

¹Information regarding purchase of copies of the documents available from EDRS can be obtained by writing to ERIC Document Reproduction Service (EDRS), Computer Microfilm International Corporation, P.O. Box 190, Arlington, VA 22210.

Borich, Gary D. Handbook for Evaluation Instruments for Preparing Educational Personnel for the Handicapped. Austin, Texas: Research and Development Center for Teacher Education, 1972 (ED 088 906).

This handbook reports and describes 51 instruments that were obtained from 11 projects and were developed by project staff to develop, assess, and change procedures for training educational personnel for the handicapped. The instruments are categorized by: the program for which the instrument was constructed, the respondents for whom the instrument is intended, and the behavioral domain that the instrument appears to measure.

Boyd, Joseph L., Jr., and Shimberg, Benjamin. Directory of Achievement Tests for Occupational Education. Princeton, New Jersey: Educational Testing Service, 1971.

This directory is an annotated listing of tests designed for the assessment of skills necessary to the automotive trades, building trades, drafting, printing, graphic arts, electronics, health, the machine trades, office, and other occupations.

Boyer, E. Gil, Simon, Anita, and Karafin, Gail, (Eds.). Measures of Maturation: An Anthology of Early Childhood Observation Instruments. Philadelphia, Pennsylvania: Research for Better Schools, Inc., 1973.

Braswell, James S. Mathematics Tests Available in the United States, (3rd ed.). Washington, D.C.: National Council of Teachers of Mathematics, Inc., 1972 (ED 067 270).

This comprehensive listing describes mathematics tests in eight categories: Arithmetic, Junior High, Batteries and General Listings K-14, High School General, Algebra, Geometry, Trigonometry, and College-Related. A directory of publishers and descriptions of a number of related evaluation services are also presented.

Buell, Kenneth, et al. An Annotated List of Tests for Spanish Speakers. Princeton, New Jersey: Educational Testing Service, 1973 (ED 079 393).

A bibliography of instruments which may be used with persons whose language is Spanish, the tests are presented in the following categories: Verbal Achievement and Aptitude/Ability; Personnel/Industrial; Intelligence; Personality; Miscellaneous.

Buros, Oscar K., (Ed.). Personality Tests and Reviews. Highland Park, New Jersey: Gryphon Press, 1970.

This book is a special supplement to the continuing series of Mental Measurements Yearbooks. It is devoted to personality tests, and contains all relevant information previously published in the first six Yearbooks as well as recent material appearing after the publication of the Sixth Mental Measurements Yearbook in 1965. It includes: a list of personality tests, each in-print test together with a bibliography of all published references to it, up to November 1968; reprints of reviews of personality tests that have appeared in the Yearbook; a directory of publishers of personality tests; and a cumulative index of titles listed in the Yearbooks and this volume.

Buros, Oscar K., (Ed.). Reading Tests and Reviews. Highland Park, New Jersey: Gryphon Press, 1968.

This is a bibliography of reading tests available as of May 1968, and reprints of test reviews from the first six Mental Measurements Yearbooks. A cumulative index to this volume and the first six Yearbooks is included.

Buros, Oscar K., (Ed.) The Seventh Mental Measurements Yearbook. Highland Park, New Jersey: Gryphon Press, 1972.

The Yearbooks, published periodically since 1932, are a comprehensive source of factual and evaluative information about tests which generally are available for open sale. The Seventh Yearbook is the most recent in the series. The two volumes contain information on 1,157 tests. In addition to descriptive information and test reviews, there are bibliographic references to studies and articles pertaining to specific instruments, reviews of publications related to measurement and evaluation, and a current directory to 243 test publishers. Not included in the Yearbooks are tests not generally available, research instruments, or measures which have not been published as separates (that is, tests included in the appendix of a book, research report, or in a journal article).

Buros, Oscar K., (Ed.). Tests in Print II (TIP II): An Index to Tests, Test Reviews, and the Literature on Specific Tests. Highland Park, New Jersey: Gryphon Press, 1974.

This reference book presents 2,467 test entries, 16,594 references on specific tests, a cumulative name index for each test covering all references in TIP II and the seven Mental Measurements Yearbooks, a directory of 493 test publishers with all tests of each publisher listed, a title index covering all tests in print and all out-of-print tests once listed in an MMY, a name index to authors of over 70,000 documents (tests, reviews, excerpts, and references) in the seven MMYs and TIP II, and a scanning index for quickly locating tests designed for a particular population.

Cattell, Raymond B., and Warburton, Frank W. Objective Personality and Motivation Tests: A Theoretical Introduction and Practical Compendium. Urbana, Illinois: University of Illinois Press, 1967.

This volume is in two parts. Part I presents a statement of essential measurement theory and develops the general personality measurement theory. Part II is a compendium comprising over 600 objective personality tests.

Chun, Ki-Taek, Cobb, Sidney, and French, John R.P., Jr. Measures for Psychological Assessment: A Guide to 3,000 Original Sources and Their Applications. Ann Arbor, Michigan: Institute for Social Research, 1975.

This volume is a compilation of annotated references to measures of mental health and related variables and the uses of these measures. The book consists of two major sections. The section on Primary References lists approximately 3,000 references to articles in journals or other publications in which measures were first described. The

Applications Section provides information on approximately 6,600 instances in which the measures cited in the first section have been used. An author index and a descriptor index are also provided.

Comrey, Andrew L., Backer, Thomas E., and Glaser, Edward M. A Sourcebook for Mental Health Measures. Los Angeles: Human Interaction Research Institute, 1973.

The Sourcebook provides detailed descriptions of over 1,100 unpublished psychological measures. The instruments are grouped into 45 mental health-related categories, including educational adjustment, family interaction, geriatrics, mental health attitudes, occupational adjustment, personality, racial attitudes, social issues, and therapeutic processes. Each of the abstracts presents identifying information followed by a brief description of the measure which includes such details as purpose, number and type of items, technical data, and source from which it can be obtained.

Cook, John J. Test Anthology: Fugitive and Standardized Tests. Madison, Wisconsin: Wisconsin State Department of Public Instruction, 1971 (ED 069 051).

This compendium of tests, behavior rating scales, and summary evaluation sheets was designed for use by special education administrators, teachers, and program planners. It contains approximately 300 entries.

Crago, Pricilla Haywood. Diagnostic Tests in Reading: An Annotated Bibliography. Albany, New York: Bureau of Pupil Testing and Advisory Services, New York State Education Department, 1970 (ED 073 426).

This is a list of 38 published diagnostic reading tests. A description, the author's comments, and a critical review abstracted from the Mental Measurements Yearbooks accompany each entry. An introduction outlining the considerations to be followed in selecting diagnostic tests and providing background for the comprehension of technical information is included.

Ehrlich, Alan and Ehrlich, Roselin. Tests in Spanish and Other Languages, English as a Second Language, and Nonverbal Tests for Bilingual Programs: An Annotated B.E.A.R.U. Bibliography. New York: Bilingual Education Applied Research Unit, May 1974 (ED 078 713).

This annotated bibliography of instruments currently available for use in the assessment of bilingual programs is a complete updating and revision of the original listing first published in January 1973.

Fagan, William T., Cooper, Charles R., and Jensen, Julie M. Measures for Research and Evaluation in the English Language Arts. Urbana, Illinois: National Council of Teachers of English, 1975.

This monograph provides descriptions and summaries of over 100 unpublished instruments for research and evaluation of student learning in the language arts. The measures are organized into eight categories: language development, listening, literature,

reading, standard English as a second language or dialect, teacher competency, writing, and miscellaneous. Information for obtaining these instruments is provided.

Farr, Roger and Anastasiow, Nicholas. Tests of Reading Readiness and Achievement: A Review and Evaluation. Newark, Delaware: International Reading Association, Inc., 1969.

This review is written for classroom teachers and other school personnel who are responsible for selecting reading readiness and elementary level reading achievement tests. The authors have reviewed and compared the five most recently used reading readiness measures and the five most recently used reading achievement tests.

Frankenburg, William K. and Camp, Bonnie W., (Eds.) Pediatric Screening Tests. Springfield, Illinois: Charles C. Thomas, 1975.

This four-part volume concerns procedures for screening children for diseases and physical handicaps. The first part consists of a discussion of the principles and criteria to be considered in selecting diseases for which to screen, and appropriate screening devices. The last three parts present specific information about diseases and screening procedures in the areas of pathophysiology, special sensory defects, and psychopathology. Chapters in these latter parts begin with a general introduction followed by reviews of specific screening tests and procedures.

French, John W. The Description of Personality Measurements in Terms of Rotated Factors. Princeton, New Jersey: Educational Testing Service, 1953.

This is a survey of factor studies of personality measurement and indicates the kinds of instruments that can measure each factor.

Goldman, Bert A. and Saunders, John L. Directory of Unpublished Experimental Measures, (Vol. I). New York: Behavioral Publications, Inc., 1974.

First in a series of indexes to experimental test instruments that are not available commercially, this Directory is based on the 1970 issues of 29 professional journals. Each test entry includes the following information: test name, purpose, description, statistics, source, and related research. The tests are grouped into 22 categories according to the purpose of the test and/or its apparent content.

Griffin, Margaret, Hibbard, Louise, and Muldoon, Kathleen. Guide to Clinical Evaluation Instruments in Reading. Bloomington, Indiana: ERIC/CRIER, 1972.

This guide to published reading tests presents specific information concerning the measures, including: author, publisher, age or grade level, subtests, forms, and administration time. In addition, notations on the availability of normative data, indexes to the Mental Measurements Yearbooks, publishers' addresses, test descriptions, and complete bibliographic references are provided.

Guthrie, P.D. Measures Pertaining to Health Education: I. Smoking. An Annotated Bibliography. Princeton, New Jersey: ERIC Clearinghouse on Tests, Measurement, and Evaluation, Educational Testing Service, 1972 (ED 060 042).

This is a bibliography of instruments that measure attitudes, behaviors, practices, knowledge, and correlates of tobacco smoking. Thirty instruments are described, including some general drug usage measures containing a number of items on smoking. General health tests, even if some items pertain to tobacco, are excluded.

Guthrie, P.D. Measures Pertaining to Health Education: II. Drugs. An Annotated Bibliography. Princeton, New Jersey: ERIC Clearinghouse on Tests, Measurement, and Evaluation, Educational Testing Service, 1972 (ED 068 569).

This is a compendium of instruments primarily concerned with drugs other than nicotine or alcohol. Ninety tests designed to assess attitudes, behaviors, practices, knowledge, and correlates of drug usage are described. Excluded are general health tests, measures of the physiologic actions of drugs, measures of the subjective effects of drugs, and standard personality measures.

Guthrie, P.D. Measures Pertaining to Health Education: III. Alcohol. An Annotated Bibliography. Princeton, New Jersey: ERIC Clearinghouse on Tests, Measurement, and Evaluation, Educational Testing Service, 1972 (ED 068 570).

This bibliography describes instruments concerned with the measurement of attitudes, behaviors, practices, knowledge and/or the correlates of alcohol usage. There are 113 instruments in the bibliography, some of which are intended for use with general populations and others are designed for use with alcohol-dependent respondents.

Hamilton, Don. Evaluation Instruments for Bilingual Education: A Revision of Tests in Use in Title VII Bilingual Education Projects. Austin, Texas: Dissemination Center for Bilingual Bicultural Education, 1973 (ED 087 818).

This annotated bibliography of tests developed by or used in Title VII projects is intended as a directory of information for use by bilingual bicultural education projects. Each annotation includes details on the applicability of the test, time required to administer it, skills assessed by the instrument, and scores yielded by its administration.

Hoepfner, Ralph (Ed.). CSE Elementary School Test Evaluations. Los Angeles: Center for the Study of Evaluation, UCLA Graduate School of Education, 1970.

This is a compendium of tests, keyed to educational objectives of elementary school education, and evaluated by measurement experts and educators. Evaluations are assigned for meaningfulness, examinee appropriateness, administrative usability, and quality of standardization. Commercial assessment, diagnostic, and prognostic measures appropriate for use in grades one through six are included.

Hoepfner, Ralph, Stern, Carolyn, and Nummedal, Susan G., (Eds.) CSE-ECRC Pre-School/Kindergarten Test Evaluations. Los Angeles, California: UCLA Graduate School of Education, School Evaluation Project, Center for the Study of Evaluation, and Early Childhood Research Center, 1971.

In this compendium of published assessment, diagnostic, and prognostic instruments keyed to educational objectives for preschool and kindergarten education, each measure is evaluated in terms of its meaningfulness, examinee appropriateness, administrative usability, and quality of standardization.

Ilg, Frances L. and Ames, Louise Bates. School Readiness-Behavior Tests Used at the Gesell Institute. New York: Harper and Row, Publishers, 1972.

The combined text and manual presents the basic educational viewpoint of the Gesell Institute and describes the tests that make up the Institute's developmental examination for young children. It explains how to give the tests, how to evaluate them, and how to use the results.

Johnson, Orval G., and Bommarito, James W. Tests and Measurements in Child Development: A Handbook. San Francisco, California: Jossey-Bass Inc., Publishers, 1971.

The Handbook provides a detailed description of over 300 instruments to assess cognitive, affective, and sensory-motor attributes. The authors' criteria for including a measure were: suitable for use with children between birth and twelve years, available to interested professionals, not commercially available, authors provide enough information for effective use, long enough to permit development of norms and technical data, and do not require heavy laboratory equipment.

Jurgens, James T. Interest Inventories, Tests, Surveys, and Scales: A Compilation. Part I of Research and Development Project in Career Education, Final Report, (Vol. III). Olympia, Washington: Washington State Coordinating Council for Occupational Education, 1973 (ED 085 536).

This list contains 66 currently available interest measures which may be used by teachers and counselors.

Knapp, Joan. An Omnibus of Measures Related to School-Based Attitudes. Princeton, New Jersey: Center for Statewide Educational Assessment, Educational Testing Service, 1972 (ED 074 071).

This work provides summaries of 16 paper-and-pencil self-report inventories that measure school-based attitudes. Some of the instruments are appropriate for children 4-8 years of age while others are for students in grades 12-14.

Knapp, Joan A. A Selection of Self-Concept Measures. Princeton, New Jersey: Center for Statewide Educational Assessment, Educational Testing Service, 1973.

The instruments described in this report were selected on the basis of their appropriateness for a normal population of school age children, and their representativeness of a variety of measurement techniques, including self-report, projective, behavior trace, and direct observation.

Lake, Dale G., Miles, Matthew B., and Earle, Ralph B., Jr. (Eds.) Measuring Human Behavior: Tools for the Assessment of Social Functioning. New York: Teachers College Press, 1973.

This reference book presents information on 84 research instruments appropriate for studying organizations, groups, and interpersonal relationships. Details on availability, variables measured, administration and scoring, history of development, reliability, validity, norms, response sets, measurement difficulties, and references are provided for each instrument. Twenty other compendia of research instruments are also described.

Lehman, Paul B. Tests and Measurements in Music. Englewood Cliffs, New Jersey: Prentice-Hall, Inc., 1968.

This includes descriptions and comments on tests of musical aptitude, achievement, performance, and appreciation. There is also an appendix which lists out-of-print music tests.

McCloy, Charles H. and Young, Norma D. Tests and Measurements in Health and Physical Education, (3rd Ed.) New York: Appleton-Century-Crofts, 1970.

Numerous tests and assessment devices in a wide variety of specific areas of health and physical education are provided and discussed in detail.

Mauser, August J. Assessing the Learning Disabled: Selected Instruments. San Rafael, California: Academic Therapy Publications, 1976.

This annotated bibliography describes more than 300 tests that may be used in the evaluation of individuals with specific learning disabilities. Tests are presented in the following areas: Intelligence; Preschool Readiness; Motor, Sensory, and Language; Reading Readiness; Diagnostic Reading; Survey Reading; Oral Reading; Diagnostic Tests of Math Abilities; Creativity; and Vocational.

Mayer, Victor J. Handbook of Unpublished Evaluation Instruments in Science Education. Columbus, Ohio: Center for Science and Math Education, Ohio State University, 1973 (ED 076 359).

This book provides brief descriptions of 156 unpublished instruments that were constructed between 1964 and 1968. The tests are classified into the following categories: Achievement in Science, Achievement in Processes and Skills of Science, Characteristics and Abilities of Students, Attitudes, Knowledge of the Nature of Science, and Professional Practices.

Pfeiffer, J. William, and Heslin, Richard. Instrumentation in Human Relations Training: A Guide to 75 Instruments with Wide Application to the Behavioral Sciences. Iowa City, Iowa: University Associates, 1973.

This book is intended as a reference resource for practitioners in the field of human relations training. Part I provides a general discussion of instrumentation in this field. Part II presents descriptions of 75 instruments applicable for groups and group-related activities. Each annotation includes information intended to facilitate decision making as to the appropriateness of the instrument in meeting specific needs: length of instrument, time estimates for administering and scoring, description of scales, specific uses, positive features of the instrument, and concerns in using the instrument.

Robinson, John P., Athanasiou, Robert, and Head, Kendra B. Measures of Occupational Attitudes and Occupational Characteristics. Ann Arbor, Michigan: Institute for Social Research, The University of Michigan, 1969.

This review and evaluation of 77 "major" empirical measures of variables related to occupations includes a copy of each of the scales evaluated.

Robinson, John P., Rusk, Jerrold G., and Head, Kendra B. Measures of Political Attitudes. Ann Arbor, Michigan: Institute for Social Research, The University of Michigan, 1968.

In this review and evaluation of 95 empirical scales designed to measure various political attitudes, some of the topics covered are: community-based political attitudes, political participation, hostility-related national attitudes, international affairs, racial and ethnic attitudes, domestic government policies, democratic principles, etc. A copy of each measure evaluated is included.

Robinson, John P. and Shaver, Phillip R. Measures of Social Psychological Attitudes: Revised Edition. Ann Arbor, Michigan: Institute for Social Research, University of Michigan, 1973.

This revised volume provides a comprehensive listing and evaluation of empirical scales for measuring such social psychological attitudes as life satisfaction and happiness, self-esteem, locus of control, alienation and anomie, dogmatism, socio-political attitudes, values, and religious attitudes.

Rosen, Pamela, (Ed.) Test Collection Bulletin. Princeton, New Jersey: Educational Testing Service, 1967 to present.

A quarterly digest of information on recent acquisitions to the Educational Testing Service Test Collection and Head Start Test Collection, the Bulletin includes announcements of forthcoming tests, publisher changes, information on scoring services and systems, new reference materials on measurement and evaluation, information on testing programs, and references to test reviews. The October Bulletin includes an index to all issues published during the year (January-October).

Savard, Jean-Guy. Analytical Bibliography of Language Tests. Quebec, Canada: Laval University Press, 1969.

This annotated bibliography covers more than 400 measures including second language tests, mother tongue tests, bilingual tests, language aptitude tests, and psychological tests. While the bulk of the text is in French, introductory and explanatory material is provided in both French and English.

Shaw, Marvin E., and Wright, Jack M. Scales for the Measurement of Attitudes. New York: McGraw-Hill Book Company, 1967.

In this collection and description of attitude scales used in research, the authors have included measures which they feel have been carefully constructed, are currently appropriate, and have at least a minimum validity and reliability. Included are scales to measure attitudes toward social and other significant institutions, ethnic and national groups, social and international issues, political and religious attitudes. Many scales are included in the book.

Simon, Anita and Boyer, E. Gil, (Eds.). Mirrors for Behavior: An Anthology of Classroom Observation Instruments. Philadelphia: Research for Better Schools, Inc., 1967-1970.

A 15-volume series concerning systems for the analysis of classroom interaction, this anthology provides an analysis of the state of the art, a compilation of 78 observation instruments, and related bibliographic references.

Solleder, Marian K. Evaluation Instruments in Health Education: 1969 Revised Edition. Washington, D.C.: American Association for Health, Physical Education, and Recreation, 1969.

This revision lists 135 tests that can be used at the elementary, junior high, senior high, or college level. Both commercially available and research instruments are described.

Straus, Murray A. Family Measurement Techniques: Abstracts of Published Instruments, 1935-1964. Minneapolis, Minnesota: University of Minnesota Press, 1969.

This is a detailed description of 319 measurement techniques developed to quantify the properties of the family or the behavior of individuals in family roles. Criteria for inclusion were: the instrument must provide a classification or numerical score; the behavior measured must refer to the subject's action or disposition to act in a family role; and the instrument must be described in a published source so that it is generally available. Whenever a scale was not contained in a published source, Doctor Straus made arrangements with its author to have a copy deposited with the National Auxiliary Publications Service so that microfiche or hard copy could be obtained easily.

Suydam, Marilyn N. Unpublished Instruments for Evaluation in Mathematics Education: An Annotated Listing. Columbus, Ohio: ERIC Information Analysis Center for Science, Mathematics, and Environmental Education, Ohio State University, 1974 (ED 086 518).

This listing includes approximately 200 unpublished instruments designed to assess mathematical instruction and reported in journals, dissertations, and ERIC documents from 1964 through 1973. The bibliography is in two parts: Part I is an annotated listing of instruments; Part II consists of a supplementary list of instruments that were not annotated because reliability and/or validity information was not provided or a copy of the instrument was not available.

Tarczan, Constance. An Educator's Guide to Psychological Tests: Descriptions and Classroom Implications. Springfield, Illinois: Charles C. Thomas Publisher, 1972.

This guide is a sourcebook of information concerning intelligence, mental age, and psychometric terminology. Approximately 60 psychological, achievement, reading, projective, social, infant, and pre-school measures are described. Information concerning the purpose and use of the test, content, administration, standardization, reliability, and validity is provided for each measure.

Tognetti, Ann. ERS Annotated Bibliography: The Evaluation of Instructional Programs. Arlington, Virginia: Educational Research Service, 1974.

This annotated bibliography is intended to assist school administrators, curriculum leaders, and others in identifying documents that might aid them in deciding whether a program should be continued, improved upon, or discontinued. The cited documents contain descriptions and analyses of strategies, techniques, problems, and issues of program evaluation as well as substantive findings or recommendations.

Walker, Deborah Klein. Socioemotional Measures for Preschool and Kindergarten Children: A Handbook. San Francisco: Jossey-Bass Inc., Publishers, 1973.

In Part I of this comprehensive handbook, an overview of socio-emotional measurement technology is presented, including discussions on the definition and classification of measures, the importance of measures, locating available measures, measurement techniques, general problems with measures for young children, and recommendations for future work. Part II provides descriptions and evaluations of 143 socioemotional measures in six areas: attitudes, general personality and emotional adjustment, interests or preferences, personality or behavior traits, self concept, and social skills or competency.

Wall, Janet and Summerlin, Lee. Standardized Science Tests: A Descriptive Listing. Washington, D.C.: National Science Teachers Association, 1973 (ED 077 706).

This is a compendium of standardized science tests, published between 1959 and 1972, available to elementary and secondary school teachers. The listing is divided into seven sections: Elementary Science Tests,

Biology Tests, Chemistry Tests, Earth Science Tests, General Science Tests, Physics Tests, and Other Science Tests.

Zytowski, Donald C., (Ed.). Contemporary Approaches to Interest Measurement. Minneapolis: University of Minnesota Press, 1973.

This anthology on interest and work value inventories begins with a presentation of basic information on the selection and use of interest inventories. Subsequent chapters provide detailed discussions of individual inventories, including: the Strong Vocational Interest Blank, the 1969 Revision of the Strong Vocational Interest Blank for Women, the Kuder Occupational Interest Survey, the Minneapolis Importance Questionnaire, the Ohio Vocational Interest Survey, and the Work Values Inventory. The final chapter contains numerous illustrative interpretations of profiles obtained for each of the inventories.

Directory of Task Inventories, (Vols. I and II). Columbus, Ohio: The Center for Vocational Education, The Ohio State University, 1974 and 1975.

These volumes serve as a directory of task inventories now being collected by the Task Inventory Exchange. They were compiled from a review of documents from state educational and employment service agencies, occupational curriculum laboratories and research coordinating units, branches of the armed forces, and other similar organizations. The following information is provided for each document cited: accession number, title, author, publication, number of pages, sponsoring agency, occupational area or job title of listed tasks, number of tasks, U.S.O.E. subject matter classification, code number, and information on availability.

Drug-Abuse Research Instrument Inventory. Cambridge, Massachusetts: Social Systems Analysts, 1973 (ED 076 611).

An annotated bibliography of instruments appropriate for use in drug-abuse research, the inventory is divided into six sections: Attitudes, Measurement of Effects of Drugs, Differentiation and Characteristics of Abusers, Access and Extent, Education and Knowledge, and Program-Related Evaluation.

Evaluation Bibliography: Parent, Child Decision Makers. Chapel Hill, North Carolina: Technical Assistance Development System, University of North Carolina, 1973 (ED 081 789).

This annotated bibliography includes tests measuring one or more of the following dimensions: language, cognition, self-help, social-affective, visual-motor, and physical health. The tests are arranged in the following age groups: 0-6 months, 6-12 months, 24-36 months, 36-48 months, and 48 months and up. Ten tests for parents are also listed.

Instruments That Measure Self-Concept. Chicago, Illinois: Chicago Board of Education (ED 079 431).

This is a listing of 54 instruments that measure self-concept. Most of the instruments included in the bibliography are appropriate for use at the elementary level.

APPENDIX C-2 - Selected References on Educational Measurement and Evaluation

The references annotated in this bibliography are sources of information regarding educational measurement. Both simple, easy-to-understand books as well as complex, theoretical volumes are included to present a broad range of possible references for educators concerned with educational measurement and related issues.

Allen, Robert M. and Allen, Sue P. Intellectual Evaluation of the Mentally Retarded Child: A Handbook. Beverly Hills, California: Western Psychological Services, 1967.

This is designed to enable the psychologist or specialist working with retarded children to select appropriate assessment procedures. Detailed descriptions of tests are provided.

Anastasi, Anne. Psychological Testing. New York: MacMillan, 1968

This text is intended to provide basic information about tests and measurement. Procedures for testing as well as a clear presentation of theoretical background information is presented.

Anderson, Scarvia B., Ball, Samuel, Murphy, Richard T., and Associates. Encyclopedia of Educational Evaluation. San Francisco: Jossey-Bass Publishers, 1975.

This is a compendium of information for professional educators concerned with program evaluation. The book is arranged like an encyclopedia. Documented articles on the main concepts and techniques in educational evaluation are presented in alphabetical order.

Barrow, Harold M. and McGee, Rosemary. A Practical Approach to Measurement in Physical Education: Second Edition. Philadelphia, Pennsylvania: Lea and Febiger, 1971.

This four-part book is designed for prospective physical education teachers in elementary and secondary schools. Part II, Evaluation of the Product, constitutes more than half the book and presents a discussion of tests in specific subject areas. Several tests are listed and described for each subject area.

Beatty, Walcott H. Improving Educational Assessment and an Inventory of Measures of Affective Behavior. Washington, D.C.: Association for Supervision and Curriculum Development, National Education Association, 1969.

The first section of this book is concerned with assessment. The second section, "Inventory of Measures of Affective Behavior," is an annotated list of measures in the following areas: attitudes, creativity, interaction, motivation, personality, readiness, self-concept, and miscellaneous.

Beggs, Donald L. and Lewis, Ernest L. Measurement and Evaluation in the Schools. Boston: Houghton Mifflin Company, 1975.

This text is an introduction to the measurement process with a special emphasis on its relation to the problems of educational evaluation.

Bloom, Benjamin S., Hastings, J. Thomas, and Madaus, George F. Handbook on Formative and Summative Evaluation of Student Learning. New York: McGraw-Hill, 1971.

This is a book about the "state of the art" of evaluating student learning. Part 1 consists of four sections dealing with the evaluation problems all teachers are likely to encounter. Part 2 consists of chapters dealing with evaluation in each of the major subject fields and levels of education.

Bronson, William H. et al. Program Evaluator's Guide. Princeton, N.J.: Educational Testing Service, 1977.

This basic manual leads the student through the steps required to develop and carry out a comprehensive plan of evaluation. There are chapters on purposes and objectives, planning and design, developing or selecting instruments, collecting and analyzing data, and reporting and applying results.

Bronson, William H. et al. Workbook on Program Evaluation. Princeton, N.J.: Educational Testing Service, 1977.

This is a recording notebook with text that serves as a programmed instruction in program evaluation.

Bronson, William H. et al. Evaluation Trainer's Guide. Princeton, N.J.: Educational Testing Service, 1977.

This detailed manual for teaching program evaluation workshops is based on the Program Evaluator's Guide.

Brown, Frederick G. Principles of Educational and Psychological Testing. New York: Holt, Rinehart and Winston, 1976.

This book is intended to provide the reader with the necessary background so (s)he can make an informed, intelligent evaluation of a test whenever the need arises. The applied aspects of testing are stressed.

Butler, Annie L., Gotts, Edward E., Quisenberry, Nancy L., and Thompson, Robert P. Literature Search and Development of an Evaluation System in Early Childhood Education. Bloomington, Indiana: Indiana University, 1971 (ED 059.783).

Part A of this report, "Behavioral Objectives," describes the procedures used in constructing behavioral objectives for five-year-old children in the psychomotor, cognitive, and affective domains. Part B, "Evaluation Instruments," provides a listing of measures appropriate for use in pre-school programs for diagnostic, screening, evaluation, or research purposes. Tests, rating scales, inventories, observational procedures, and technical assessment procedures are among the types of instruments that are included.

Cattell, Raymond B., and Warburton, Frank W. Objective Personality and Motivation Tests: A Theoretical Introduction and Practical Compendium. Urbana, Illinois: University of Illinois Press, 1967.

This volume is in two parts. Part I presents a statement of essential measurement theory and develops the general personality measurement theory. Part II is a compendium of more than 600 objective personality tests.

Chase, Clinton I. Measurement for Educational Evaluation. Reading, Massachusetts: Addison-Wesley, 1974.

This book presents some basic generalizations from which educators can assess their own problems. General guidelines for building and evaluating tests are presented. Emphasis is on features to note in various kinds of tests.

Coller, Alan R. The Assessment of Self-Concept in Early Childhood Education. Urbana, Illinois: ERIC Clearinghouse on Early Childhood Education, July 1971.

This is a state-of-the-art paper that includes descriptions and discussions of numerous measures of self-concept.

Coller, Alan R. Systems for Observing Parent-Child Interaction. Urbana, Illinois: ERIC Clearinghouse on Early Childhood Education, June 1972.

This is a state-of-the-art paper that includes descriptions and discussions of systems useful for the observation of behaviors that occur between parents and their children.

Coller, Alan R. Systems for the Observation of Classroom Behavior in Early Childhood Education. Urbana, Illinois: ERIC Clearinghouse on Early Childhood Education, April 1972.

This book describes systems and instruments used to systematically observe early childhood classroom behavior.

Cronbach, Lee J. Essentials of Psychological Testing. New York: Harper and Row, 1970.

This is a basic text on various kinds of tests with explicit definitions of their respective purposes. Issues related to testing are also carefully considered. This book is more useful for test selection and interpretation than for test construction.

DiLeo, Joseph H. Children's Drawings as Diagnostic Aids. New York: Brunner/Mazel Publishers, 1973.

Part I of this volume provides an overview of the origin and development of representational drawings, their reflection of cognition and neurological impairment, and their use as a projective technique. The remaining three parts focus on specific applications of the technique including family drawings and the indicators of relationships in the nuclear, incomplete, and extended family; the drawings of handicapped children, the effects of sensory and motor impairment, and what the drawing reveals of the child's disability; and the value of drawings in the early detection of minimal brain dysfunction, dyslexia, mental retardation, inborn errors of metabolism, and emotional problems affecting learning.

DiLeo, Joseph H. Young Children and Their Drawings. New York: Brunner/Mazel Publishers, 1970.

This volume represents a study of the drawings of normal children from thirteen months to six years old and provides a discussion of the use of drawings as a diagnostic aid.

Ebel, Robert L. Essentials of Educational Measurement. Englewood Cliffs, New Jersey: Prentice-Hall, 1965.

This book concerns the practical methods and procedures of educational measurement. The author describes how various types of tests can be used effectively in dealing with current problems and controversies in education, and provides thorough grounding in methods of testing and evaluation.

Egner, Ann N., Burdett, Carol S., and Fox, Wayne L. Observing and Measuring Classroom Behaviors. Austin, Texas: Austin Writers Group, 1972.

This book describes a variety of techniques that teachers can use to more precisely observe and measure classroom behavior. Types of classroom behavior covered include language, arithmetic, social, and teacher.

Gronlund, Norman E. Measurement and Evaluation in Teaching. New York: Macmillan, 1971.

This volume introduces the educator to the principles and procedures of measurement and evaluation that are essential for effective teaching. Part I describes the roles of objectives, validity, and reliability in the evaluation process. Part II describes procedures for constructing all types of classroom tests. Part III introduces the reader to selected standardized tests of achievement and scholastic aptitude, stressing principles for selecting tests wisely. Part IV describes construction, selection, and use of evaluation procedures to appraise learning outcomes in nontest areas. Part V discusses the role of evaluation in improving learning, marking, and reporting.

Lambert, Nadine, Wilcox, Margaret R., and Gleason, W. Preston. The Educationally Retarded Child: Comprehensive Assessment and Planning for Slow Learners and the Educable Mentally Retarded. New York: Grune and Stratton, 1975.

This book summarizes the major investigations of the assessment of retardation in culturally different children, emphasizing the limitations as well as the usefulness of some of the standard assessment tools. Suggestions and examples are provided to help the psychologist and educator employ some of the more recent assessment methods.

Mager, Robert F. Preparing Instructional Objectives. San Francisco: Fearon, 1962.

This volume contains specific directions for the writing of objectives for instructional purposes.

Martinson, Ruth A. The Identification of the Gifted and Talented. Reston, Virginia: The Council for Exceptional Children, June 1974.

This instructional syllabus presents a rationale for the identification of gifted students, suggests appropriate identification procedures, and gives examples of materials used by various school districts.

Mattick, Ilse and Parkins, Frances J. Guidelines for Observation and Assessment: An Approach to Evaluating the Learning Environment of a Day Care Center. Washington, D.C.: The Day Care and Child Development Council of America, Inc., 1972.

An observation scheme for evaluating the learning environment of day care centers is presented and described.

Miller, Delbert C. Handbook of Research Design and Social Measurement, Second Edition. New York: David McKay Company, Inc., 1970.

This handbook for social researchers covers a wide range of practical information on every phase of research design and operation. Thirty-five instruments in a variety of social domains are included and described. Also, descriptions of and references to a number of other social measures are presented.

Myers, Elizabeth S., Ball, Helen H., and Crutchfield, Marjorie. The Kindergarten Teacher's Handbook. Los Angeles: Gramercy Press, 1973.

This book presents a program designed to provide the teacher with a means of assessing kindergarten children, and individualizing instruction accordingly. An introduction to and general discussion of the program is provided in Section I. Assessment procedures which will yield a sample of each child's functioning in visual motor integration, visual memory, fine motor and manipulative skill, language, gross motor control, body concept, auditory discrimination, and logical thinking processes are described in Section II. Suggested instructional activities and exercises in concept discovery are presented in Section III and IV respectively. Finally, Section V enables the teacher to refresh her classroom management skills.

Popham, W.J. Criterion-Referenced Measurement: An Introduction. Englewood Cliffs, New Jersey: Educational Technology Publications, Inc., 1971. (ED 053 200)

At the 1970 meeting of the American Educational Research Association, a symposium, jointly sponsored with the National Council on Measurement in Education, was presented on the topic "Criterion-Referenced Measurement: Emerging Issues." This text contains the following papers from the symposium: "Instructional Technology and the Measurement of Learning Outcomes: Some Questions" (Robert Glaser); "Implications of Criterion-Referenced Measurement" (W. James Popham and T.R. Husek); "A Criterion-Referenced Test" (Robert Glaser); "The Applicability of Criterion-Referenced Measurement by Content Area and Level" (Alfred D. Garvin); "Evaluation Aspects of Criterion-Referenced Measures" (Richard C. Cox); and, "Indices of Adequacy for Criterion-Referenced Test Items" (W. James Popham). Selected references and biographical data about the authors are included.

Popham, W. James. Educational Evaluation. Englewood Cliffs, New Jersey: Prentice-Hall, 1975.

This book offers an in-depth examination of the technical advances developed in the past decade and explains important issues, such as criterion-referenced testing, assessment of affect, the role of instructional objectives in evaluation, designing evaluation studies for meaningful results, contemporary evaluation models, teacher evaluation, and many more practical issues.

Sattler, Jerome M. Assessment of Children's Intelligence. Philadelphia: W.B. Sanders Company, 1974.

This text is intended for those who have a background in tests and measurement. It is comprised of six sections: Introduction and General Considerations; Administering Individual Intelligence Tests; Stanford-Binet Intelligence Scale; WISC, WPPSI, and Other Tests; Diagnostic Applications; and Psychological Reports and Consultation.

Scott, M. Gladys and French, Esther. Measurement and Evaluation in Physical Education. Dubuque, Iowa: Wm. C. Brown Company Publishers, 1959.

This volume discusses the uses of measurement and evaluation in physical education and provides a background for test construction, selection, and use. Detailed information is presented for a number of tests.

Thorndike, Robert L. (Ed.) Educational Measurement. Washington, D.C.: American Council on Education, 1970.

This volume presents a complete discussion on issues in measurement and evaluation. The various chapters are written by those with special expertise in the topics of their respective chapters.

Tollefson, Nona F. Testing and Assessment in Elementary School Guidance Programs. Boston: Houghton Mifflin Company, 1975.

This monograph reviews the rationale and purposes underlying testing and assessment of children in the elementary school. It is organized in terms of assessment of the cognitive, affective, and psychomotor domains. Specific instruments within each area are described and evaluated.

Tuckman, Bruce W. Measuring Educational Outcomes: Fundamentals of Testing. New York: Harcourt Brace Jovanovich, 1975.

This is a four-part book intended to help educators to (a) construct their own tests, (b) judge and evaluate the quality of tests, (c) choose from among published tests, and (d) use and interpret test results.

Whybrew, William E. Measurement and Evaluation in Music: Second Edition. Dubuque, Iowa: Wm. C. Brown Company, Publishers, 1971.

This is a guide to the selection, use, and interpretation of achievement and aptitude tests in music. Some of the topics covered are: the psychological bases of musical aptitude tests, problems in music testing, analyses of existing measures of musical aptitude and achievement, construction of classroom tests, and course grading.

Wisland, Milton V. (Ed.) Psychoeducational Diagnosis of Exceptional Children. Springfield, Illinois: Charles C. Thomas, 1974.

This book provides comprehensive coverage of the problems of the diagnostic processes associated with exceptional children. Three introductory chapters present background for evaluating and interpreting psychoeducational instruments. The last seven chapters deal with specific assessment procedures associated with various exceptional children including the mentally handicapped, blind and partially sighted, deaf and hard of hearing, speech handicapped, physically and neurologically impaired, socially and emotionally handicapped, and the gifted.

Wylie, Ruth C. The Self Concept: A Critical Survey of Pertinent Research Literature. Lincoln, Nebraska: University of Nebraska Press, 1961.

Included in the book are discussions of the various methods of assessing self concept and descriptions of research instruments devised to assess self concept.

Wylie, Ruth C. The Self Concept: A Review of Methodological Considerations and Measuring Instruments. (Vol. 1) Lincoln, Nebraska: University of Nebraska Press, 1974.

This volume is concerned with a conceptual treatment of methodology relevant to self-concept research and an evaluation of extant research designs, procedures, and measurement techniques.

APPENDIX D - Glossary of Selected Terms

Academic Aptitude is the combined effect of innate and acquired abilities that are essential for school learning. It is generally estimated by assessing the abilities necessary for doing academic work.

Achievement Age is the average performance level of a particular age group and is expressed in terms of chronological age.

Achievement Test is one that measures the extent to which a person has acquired and can demonstrate certain skills and proficiencies that are usually a result of explicit or implicit instruction.

Administrative Order (NJAC) is a written directive ordering specific corrective action by a district that has shown insufficient educational progress toward meeting goals and standards within a reasonable period of time.

Appropriateness is the extent to which the test as a whole and the test items represent what they were intended to measure. This can be determined by comparing the items with the test specifications. To determine appropriateness for a specific program, the test specifications and the test items should be compared to the goals, objectives, and content of the program.

Approval (NJAC) is the official classification by the state Department of Education certifying that the school and district comply with prescribed standards, pursuant to law and regulation.

Aptitude is a combination of abilities known or believed to be related to a person's ability to learn. Measurements of aptitude are used to predict future general academic performance as well as performance in

specific areas, such as verbal, mathematical, mechanical, and numerical aptitude.

Arithmetic Mean is a measure of central tendency derived by summing a group of scores and then dividing by the number of scores in the group. It is commonly referred to as the "mean score" or "average."

Articulation (NJAC) means continuity, consistency, and interdependence in the curricular offering of the successive division of the school system.

Assessment (NJAC) means a written analysis of the current status of an educational system in terms of achieving its goals and objectives.

Assessment Instrument is any procedure that gathers information on a person, object, situation, etc., and is used for evaluation.

Assessment Objectives are a combination of standards and statements of conditions that would exist if a defined goal were being attained; these are used to identify discrepancies between current status and desired goals.

Average is a general term referring to the measures of central tendency.

These measures are of three types: mean, median, mode. The term is most commonly used with reference to the arithmetic mean (see mean, median, mode in this appendix).

Battery is a group of tests administered together in a predefined order to the same sample population so that the results are comparable. The results give a more comprehensive indication of overall performance. Generally, several areas are tested, such as reading, mathematics, and social studies.

Behavioral Rating Scale is an instrument used in evaluating the frequency and nature of particular behaviors. This scale can be greatly

influenced by subjective factors and should therefore be tested for reliability.

Ceiling is the highest level of performance that can be measured by a particular test.

Checklist is a list of behaviors, events, or characteristics, each of which can be recorded by an observer as either "yes," indicating the behavior has occurred satisfactorily, or "no," indicating that it has not.

Chronological Age is an individual's actual age expressed in years and months.

Classification (NJAC) is a process through which districts and schools are designated as approved, conditionally approved or unapproved, pursuant to law and regulation.

Completion Item is a test item type which requires filling in a blank with a word or phrase.

Conditionally Approved (NJAC) means that a school district with specific deficiencies is granted a prescribed period of time in which to plan and implement remediation pursuant to law and regulation.

Content Outline is a specification of instructional goals and the emphasis for each goal for a particular subject or curriculum area. This outline can be used to define the content of a test in order to assess how well the goals have been reached.

Correlation is the relationship between two scores or measures; it is the tendency of one score to rise or decline when another score rises or declines for the same sample. Although scores may vary in the same way with respect to each other, this does not mean that a change in one score causes the change in the other.

Correlation Coefficient is a measure of the degree of relationship between two scores or measures. A correlation coefficient can range from +1 to -1 with .00 at the midpoint. A correlation coefficient of .00 indicates a complete absence of relationship while a correlation coefficient of +1 or -1 indicates a perfect positive or a perfect negative relationship.

Criterion is a standard or particular level of proficiency by which performance can be compared and evaluated.

Criterion-Referenced Test (Measure) is an instrument used to ascertain whether a learner has achieved a predefined set of objectives at their specified levels of proficiency. The learner's abilities are compared to the criterion and not to other individuals tested.

Criterion Score is an index of the level of the student's ability. (In ETS item analysis, a distribution of criterion scores has a mean of 13 and a standard deviation of 4).

Culture Fair Test is a test that is specifically designed to assure all candidates an equal probability of success by including content material common or uncommon to all individuals regardless of their background.

Data Collection involves all the procedures necessary for the accumulation, summarization, and reporting of data instrumental in making evaluative decisions.

Decile is any one of the nine percentile points that divides a distribution of scores into ten parts, each containing ten percent of the total distribution.

Demography is the composition of a human population for a given area in terms of occupation, education, sex, ethnic background, etc.

Deviation is the degree of difference between a particular score and a reference score such as the mean, the norm, or the score on some other test. For example, where X equals an individual score and $\bar{X}$ equals the mean of a distribution of scores that includes X , $X - \bar{X}$ equals the deviation of that score from its mean.

Diagnostic Test is a test designed to locate particular problem areas by dividing a larger body of information, indicators of ability or skills, into its component parts or sub-objectives. The problem area can be determined by observing the sub-objective(s) with which the individual had difficulty. These tests are generally used in mathematics and reading.

Difficulty Index is a measure that classifies an item as easy or hard based on the percentage of students from a specified group who answer the item correctly.

Discriminability Index is a measure that reflects the extent to which an item differentiates between students who possess a little or much of a certain skill or proficiency.

Distractor is any of the incorrect options in a multiple-choice item or task.

Distribution (frequency distribution) is a table indicating the number of individuals obtaining each score within a range of obtained scores for a particular test or task.

Domain Referencing is the systematic constructing of test items or tasks to satisfy every possible condition that could exist within the particular objective the test or task is designed to measure.

Educational Plan (NJAC) is a comprehensive procedure that enables participants working together to define and set goals, to appraise

progress through assessment, to analyze problems, to plan for and to implement a program for improvement, and to evaluate progress toward achieving goals, objectives, and standards.

Equating is the process of transforming obtained scores from several different groups to a standard or common scale so that the scores can be compared and evaluated with reference to a norm group.

Essay-type Item is a test question that allows the test takers to organize, structure and compose their own answers. The item is graded according to its relevant content and the thoroughness of the response. The limitation of this item type is that grading relies on the subjective judgment of the scorer.

Evaluation (NJAC) refers to the procedures used to determine the value and success of programs, projects, techniques, and materials in relation to the achievement of goals, objectives, and standards, i.e., making judgments based upon the data gathered.

Evaluation Design means the explicitly defined procedures used in the collection of data and in evaluation.

Experimental Test is a separately timed group of items administered as a section of a test or as an individual test and is conducted to provide information about the effectiveness of the test as a whole rather than about individual items.

Extrapolation is the process of estimating the value of a variable that is beyond the range of the available data. The estimation is based on what is observed and through inference extending it to the variable in question. Since such values are rather arbitrary, they have limited meaning.

Face Validity is the acceptability of a test, based on what it appears to

measure, as an indicator of what it proposes to measure.

Factor refers to a hypothetical trait or ability responsible for performance in a particular area. Usually, one factor influences performance in more than one area, e.g., mathematical and verbal aptitude, thus causing performance on tests in the two areas to be correlated. A factor may be identified by factor analysis.

Factor Analysis refers to any of the several methods of analyzing intercorrelations among tests assessing different abilities, and attributing the intercorrelations to certain factors assumed to be influencing performance on the tests.

Faking is the tendency, on an affective measure, to give answers that are considered socially desirable or undesirable, depending upon the impression the test taker wishes to impart.

Final Form is an assembled test, containing items of a specified type and range of difficulty, used in a regular administration for data collection.

Forced-choice Item is any item that limits the response to the selection of given alternatives.

Formative Evaluation is the process of collecting evidence and information during the construction and implementation of a new curriculum and using that information to assess the effectiveness of the curriculum and correct its weak points.

Goals (NJAC) means a written statement of educational aspirations for learner achievement and the educational process stated in general terms.

Goal Statements are generalizations about the intended results of a curriculum. Goal statements can be expressed in two ways:

- (1) Outcome goals (ends) which state what a curriculum is intended to accomplish;

- (2) Process goals (means) which state how the ends are to be accomplished and are usually arrived at by the administrative staff of a program.

Grade Equivalent is the grade level for which a given level of performance (score) on a test is the real or estimated average. Assuming a ten-month school year, performance can be expressed in terms of the year and month (e.g., 6.7).

Grade Norm is the average score achieved by typical pupils at a given grade level.

Group Test is a test that can be administered to several individuals at one time and requires only one examiner.

Individual Test is a test that can be administered to only one person at a time with one examiner.

Intelligence Quotient (IQ) was originally expressed as an individual's mental age divided by his/her chronological age (MA/CA), multiplied by 100. Since the same IQ at different age levels had different meanings, this was later changed to the deviation IQ which is based on a person's score relative to the average score for persons in his/her own age group.

Interpretability is an indication of the meaning of test scores and how their meaning is determined. Test interpretation is either norm referenced or criterion referenced.

Inventory Test is an achievement test designed to thoroughly assess an individual's knowledge or skill obtained from specific instruction or training.

Item Analysis is a statistical description that indicates how a given item functioned on a particular test for a specific test administration.

This analysis includes the difficulty of the item, its discriminating power and, if the item is multiple choice, the attractiveness of the options.

Kuder-Richardson Formulas are methods of estimating the reliability of a test based on inter-item consistency. These formulas should not be used for tests that measure the rate of performance or for tests whose items cannot be scored as either right or wrong.

Likert Scale is a method of evaluating attitudes. The possible responses to each question consist of a series of statements ranging from strongly agrees to strongly disagrees, with points between the two indicating the strength of agreement or disagreement. The respondent is to circle one choice. A mid point indicating neutrality is sometimes omitted.

Matching Item is an item type that requires pairing a given entry from one list with the corresponding entry on another list.

Median is the point in a distribution of scores below which 50 percent of the scores lie. It divides a distribution of scores into two equal parts.

Mental Ability is what is traditionally called intelligence. It consists of verbal fluency and comprehension, quantitative ability, reasoning, and figural comprehension that are influenced by an individual's learning environment which reinforces or requires them.

Mental Age is the age for which a given level of performance (score) is average or normal. For example, if a five-year-old child performs at a level that is average for seven-year-olds, then that child's mental age is seven.

Mode is the most frequently occurring score in a distribution of scores.

Multiple-choice Item is a test question that has several plausible options proposed as answers, only one of which is the correct answer or the key.

The incorrect options are referred to as distractors.

Needs is a measure of the discrepancy between goals, objectives, and standards and the current status.

Needs Assessment is an evaluation to determine whether the present educational program is achieving the intended results.

Non-verbal Test is a test that does not require the use of words in the items or in the responses to the items. It does not require reading or writing; however, oral instructions are frequently used in administering the test.

Normal Distribution is a theoretical distribution of events or scores based on chance occurrences. The graph of the distribution is bell shaped and symmetrical on both sides of the mean. There is an equal number of events above and below the mean. The events are clustered near the mean, decreasing in frequency as you go to the left or the right of the mean.

Norming is the setting up of a situation to collect data for providing descriptive statistics for a test. The test is administered to a well-defined group to be used as a reference. It provides the means for interpreting the test results of other individuals who fit the description of the group and take the same test later under similar conditions.

Norming Group is a well-defined sample of people on which norm or comparison scores are obtained as part of the process of test standardization.

Norm-referenced Test is a type of test that involves interpretation of the test result through comparison with a referent or norming group.

Norms are statistics that are used as a point of reference and provide information for the interpretation of test scores. Norms are based

on the performance of those in the standardization group for the test. The most common types of norms are percentiles, grade equivalents, and stanines.

Objective (NJAC) is a written statement of the intended outcome of a specific educational process.

Omits are the number of examinees who skipped a particular test item and are assumed to have considered it since they answered at least one item after the omitted item. If there is not at least one answered item after an omitted item, the item is generally considered "not reached" rather than "omitted."

Omnibus Test is a single test that measures performance related to several mental operations. The items are in a single sequence and not grouped according to the operation tested. The test yields a single score of overall performance rather than performance scores for each operation tested.

Parallel Forms of a Test are two or more forms of tests constructed to match as closely as possible the same content and statistical specifications. This is to provide the same kind of measurement at different administrations.

Parallel Items Agreement is the extent to which two items assumed to measure the same objective actually do. The level of agreement is indicated by a correlation.

Percentile is any point or score in a distribution below which a certain percentage of cases fall. For example, the 35th percentile is the point below which 35 percent of the cases fall.

Percentile Band is the cumulative percentage of scores that falls one standard error of measurement above or below a given score.

Percentile Rank describes the relative standing of a raw score in a sequence of scores. It tells what percent of the test takers scored lower and what percent scored higher.

Percent Pass (P+) is the percentage of people who answer an item correctly on a given test.

Performance Test is a test that usually involves the manipulation of objects to resolve defined problems, and is frequently an indication of the ability to analyze. The term also refers to a test that obtains a work sample, and is identical with the behavior about which information is desired.

Personality Test is a test designed to measure the affective characteristics of an individual, such as motivation, attitude, and emotions. This is viewed separately from an individual's potentials or abilities.

Population is the total number of events, observations, or people who fit certain specified criteria.

Power Test is a test administered without time limitations in order to optimize examinee performance.

Practice Effect is the effect on test performance that is attributable to previous experience with the same or a similar test. The practice effect is generally greatest when the length of time between testings is short and the initial test taking represents a rather novel experience for the test takers.

Pretest is a subsection of a larger test or an entire test comprised of untried items. It is administered to determine the effectiveness of the items before using them in a final form.

Proficiency or Expectancy Level is the anticipated level of acquisition of a skill or ability for a given individual or group because of variables

that classify the individual(s) as part of a certain norm group, e.g., age, sex, occupation, socioeconomic status, etc., or because of participation in a certain program or curriculum with defined instructional objectives.

Proficiency Test is a test designed to assess the extent to which a person has acquired the skill or ability stated as an objective of a course of instruction or training. This is also referred to as a Criterion-referenced test.

Profile is a graphic presentation of an individual's or a group's results on several tests after the scores have been transformed into a standard scale to allow for comparisons. It helps determine strengths and weaknesses.

Prognosis (Prognostic) Test is a test designed to predict future performance in a specific area.

Projective Test is a test designed to assess personality characteristics. The examinee is presented with an ambiguous stimulus, such as an ink-blot or a picture, and is asked to interpret the stimulus, or supply it with context or meaning. It is assumed that the subject projects or manifests his/her personality characteristics in the free response.

Quartile is one of three points that divide a distribution into four equal portions. The lower quartile (Q_1), or 25th percentile, sets off the lower 25 percent; the middle quartile (Q_2) equals the 50th percentile, and divides the second fourth of the cases from the third; and the third quartile (Q_3), or 75th percentile, sets off the top 25 percent.

Random Sample is the selection of members of a population in such a way that each member has an equal chance of being selected. The intention is that the sample, when selected in this fashion, will adequately

represent the population.

Range is the difference between the highest and lowest score on a test by a group of test takers.

Raw Score is the initial quantitative result obtained in scoring a test. This score can be converted to a derived score, like a standard score or percentile.

Readiness Test is a test designed to ascertain whether an individual has reached a stage of development considered prerequisite to the learning of new skills.

Recall Item is a type of test question that requires the examinee to respond correctly to recollection or memory.

Recognition Item is a type of test question that presents two or more options as possible answers to a question. The examinee need only recognize the correct answer from the available alternatives.

Regression Effect is the tendency of students who achieve extremely high or low scores on a test to score closer to the mean on a subsequent administration of the same test.

Reliability is the extent to which a test is consistent in measuring what it does measure. For example, it should arrive at the same or very similar result on separate administrations for an individual or group if it is a highly reliable measure.

Alternate-form Reliability is a measure of the extent to which parallel forms of a test arrive at the same result.

Internal-consistency Reliability is a measure of the extent to which the items on a test measure the same abilities.

Kuder-Richardson Formula (see Kuder-Richardson Formula).

Split-half Reliability is a method of estimating the reliability of a

test by correlating scores on half of the test with the scores on the other half.

Test-retest Reliability is a method of estimating the reliability of a test by correlating the scores of a group on the first and second administrations of the test.

Reliability Coefficient is the correlation coefficient between two forms of a test, the scores on two administrations, or between halves of the same test.

Remediation is intended as an educational treatment or process to compensate for a major deficiency as indicated through needs assessment.

Representative Sample is a group of events or people selected from a population that adequately portrays the characteristics of the population. Information derived through the use of this type of sample may be used to describe the population.

Scale is a graded continuum used to classify objects or events according to some measured characteristic of it.

Scaled Score is the score obtained after the raw score is converted to a standard scale. The purpose is to make performance as independent as possible from the particular test and to facilitate the comparability of test results.

Score Scaling is the process of converting raw scores to scaled scores by setting up the standard reference scale and applying the conversion factor to the raw scores.

Semantic Differential is a scale used in measuring attitudes about a particular object. The scale uses lists of paired adjectives. Each adjective is paired with its opposite and laterally separated by seven short blank lines. Each position indicates the degree of association

with the listed adjective.

Short-answer Item is a test question that allows for only one correct response and requires recall of that response. Since there is only one correct answer this type of item can be objectively graded.

Skewness is the tendency of a distribution to have more scores at the upper or lower ends of the scale. Unlike a normal distribution, the scores are not balanced proportionately about the mean. A distribution with the bulk of the scores at the upper end of the scale is said to be negatively skewed; a distribution with the bulk of the scores at the lower end of the scale is said to be positively skewed.

Social Desirability Response Bias is the tendency for a test taker to respond in a manner that is characteristic of his/her impression of what is considered normal or acceptable.

Spearman-Brown Formula defines the relationship between the reliability of a test and its length. The reliability of a test for any given length can be estimated from a known reliability at a specified length.

Speededness is an indication of whether the examinees were able to complete the test in the time allotted. If a test is administered within an insufficient amount of time, not all test takers will complete the test and this will affect the test scores.

Spiralling is the arrangement and distribution of different forms of a test in repeated cycles to insure that the testing of various subpopulations is similar to the testing of random samples.

Standard Deviation (S.D.) is a measure of the dispersion of scores about the mean. The smaller the standard deviation, the more scores tend

to cluster about the mean. For a normal distribution approximately 68 percent of the scores fall 1 S.D. above and below the mean.

Standard Error of Measurement is the estimate of the amount of error to be expected in a test score. It is used to indicate the number of points an obtained score may vary from its true score.

Standards (NJAC) means the procedures and stated levels of proficiency used in determining the extent to which goals and objectives are being met.

Standard Score is a general term applied to a transformed raw score. The transformation of a raw score to a standard score facilitates interpretation and comparability of scores. For example, the most frequently used scaled score is the z-score which is an expression of the deviation of a score from the mean score of the group in relation to the standard deviation of the scores of the group.

Standardized Test is a test that is administered with prescribed directions and scored according to specified rules. This type of test provides a systematic sample of individual performance. Also, its results may be interpreted in reference to normative information; however, criterion-referenced tests may also be standardized. Usually the items for a standardized test are chosen on the basis of experimental evaluation. Typically such tests are published commercially.

Stanine is a point on a standard score scale consisting of 9 points with a mean of 5 and a standard deviation of 2.

Structured Item is any type of short answer test question that requires the examinee to recognize the correct answer from the list of proposed answers.

Summative Evaluation is the type of evaluation that takes place after a

course of instruction to assess the effectiveness of the program and/or evaluate the progress of the student.

Survey Test is a test that assesses general achievement in a particular area. Usually the test is intended to assess group status rather than individual performance.

T-score is a standard score scale with a mean of 50 and a standard deviation of 10.

Taxonomy is the development of principles of classification relevant to a particular body of knowledge.

Teaching Staff Members (NJAC) means all teachers, principals, assistant principals, vice principals, superintendents, assistant superintendents, school nurses and other employees who are in positions that require them to hold appropriate certificates issued by the board of examiners, serving in any school district or under any board of education.

Test Analysis is a statistical and analytical report that indicates how a test functioned when it was administered. The report includes such things as distributions of the raw and scaled scores, descriptive statistics, indices of reliability and speededness of the test, and the frequency of occurrence of chance scores.

Test-retest Reliability Coefficient is the coefficient of correlation between the two sets of obtained scores on a test administered twice to a group, with only a short interval between administrations.

True-false Item is a type of question in which a short statement is presented and the examinee is asked to determine whether it is true or false.

True Score is a hypothetical test score that is entirely free of error.

Since all tests involve some error they can never arrive at a true

score. The true score can be viewed as the average of an individual's infinite number of scores on the same test, assuming there are no practice effects or improved performance due to acquisition of new knowledge.

Unapproved (NJAC) means that a school or district has failed to show sufficient progress in implementing goals, objectives, and standards, pursuant to law and regulation.

Unstructured Item is a type of short-answer test question that requires a word, phrase, or number response. The item is similar to a completion item but is usually stated in the form of a question without the blank line, such as: Who is the author of HERZOG?

Usability is the extent to which a test is practical or suitable for its intended use.

Validity is the extent to which a measure actually measures what it is intended to measure.

Concurrent Validity is the extent to which two measures of the same thing are in agreement.

Construct Validity is the extent to which a measure, constructed according to the theory of the object of assessment, measures what it is intended to measure.

Content Validity is the extent to which a measure contains a balanced and representative sampling of the universe of the content it is intended to measure. It can be determined by comparing the content of the test with the content of the curriculum, instructional materials, and statements of instructional goals.

Criterion Validity is the extent to which a group proficient in what is being measured scores higher than a group not experienced or

proficient in what is being measured.

Face Validity is the extent to which a measure appears to measure what it is intended to measure as determined by specialists in the particular area.

Predictive Validity is the extent to which a measure indicates future outcome in an area as indicated by correlations between performance on the measure and future criterion measures.

Variance is a measure of dispersion of a set of scores about the mean of the distribution.

Weighting is assigning a quantitative value to the various subsections of a measure so that the total score or analysis of the total picture reflects that value.

APPENDIX E - Potential Sources of Information and/or
Assistance in Assessment and Evaluation

I. County Office

The county office, in connection with T & E, has responsibility for:

- (1) Assisting districts in all phases of their efforts and monitoring progress of the total educational planning process and improvement programs of local districts.
- (2) Reviewing each district's budget for adequacy with regard to long- and short-range objectives.
- (3) Conducting on-site evaluation visits.
- (4) Approval of district plans and reports.
- (5) Continuing regulatory responsibility including:
 - Certification
 - Budget preparation
 - Building facilities
 - Pupil transportation contract approvals
- (6) Serving as representative of the Commissioner in a variety of educational capacities in the county.

The county offices are:

<u>COUNTY</u>	<u>ADDRESS</u>	<u>PHONE</u>
Atlantic	State Department of Education Atlantic County Office 1200 Harding Highway Mays Landing, NJ 08330	(609) 625-2101
Bergen	State Department of Education Bergen County Office 304 Valley Boulevard Wood-Ridge, NJ 07075	(201) 935-2990
Burlington	State Department of Education Burlington County Office 3 Union Street Mount Holly, NJ 08060	(609) 267-3300

County Offices (continued)

<u>COUNTY</u>	<u>ADDRESS</u>	<u>PHONE</u>
Camden	State Department of Education Camden County Office 2276 North 43rd Street Pennsauker, NJ 08110	(609) 757-8987
Cape May	State Department of Education Cape May County Office Corner Mechanic and Boyd Streets Cape May Court House, NJ 08210	(609) 465-7542
Cumberland	State Department of Education Cumberland County Office 19 Landis Avenue Bridgeton, NJ 08302	(609) 451-8000
Essex	State Department of Education Essex County Office 90 Washington Street East Orange, NJ 07017	(201) 673-3900
Gloucester	State Department of Education Gloucester County Office Tanyard and Salina Roads Sewell, NJ 08080	(609) 468-6500
Hudson	State Department of Education Hudson County Office 595 Newark Avenue Jersey City, NJ 07306	(201) 792-3737
Hunterdon	State Department of Education Hunterdon County Office P.O. Box 87, Route 12 Flemington, NJ 08822	(201) 782-2424
Mercer	State Department of Education Mercer County Office 640 South Broad Street Trenton, NJ 08607	(609) 989-6677
Middlesex	State Department of Education Middlesex County Office 97 Bayard Street New Brunswick, NJ 08901	(201) 246-6490

County Offices (continued)

<u>COUNTY</u>	<u>ADDRESS</u>	<u>PHONE</u>
Monmouth	State Department of Education Monmouth County Office Campbell Court and Highway #9 Freehold, NJ 07728	(201) 431-7810
Morris	State Department of Education Morris County Office Court House Morristown, NJ 07960	(201) 285-6151
Ocean	State Department of Education Ocean County Office 212 Washington Street Toms River, NJ 08753	(201) 244-2121
Passaic	State Department of Education Passaic County Office 519 Ringwood Avenue Pompton Lakes, NJ 07442	(201) 839-7070
Salem	State Department of Education Salem County Office Salem-Woodstown Road P.O. Box 98 Woodstown, NJ 08098	(609) 769-2700
Somerset	State Department of Education Somerset County Office 32 Grove Street Somerville, NJ 08876	(201) 725-4700
Sussex	State Department of Education Sussex County Office 18 Church Street Newton, NJ 07860	(201) 383-2521
Union	State Department of Education Union County Office 300 North Avenue, East Westfield, NJ 07090	(210) 233-9310
Warren	State Department of Education Warren County Office 413 Second Street Belvidere, NJ 07823	(201) 475-5361

II. EIC Services

A primary point of contact to a technical assistance system which branches out to include the nation as a whole is the Educational Improvement Center (EIC), which serves different geographical areas of the State.

EICs are equipped with comprehensive educational information storage and retrieval facilities, making possible rapid response to inquiries on a broad range of topics dealing with what's new, and what's tried and proved in education today, coming from within and outside of the State of New Jersey.

They can provide personalized technical consultation services on numerous educational topics. In so doing, EIC can assist in locating, modifying or designing programs needed by local districts for use in school improvement, ranging from models for goal setting to assessment and evaluation.

They can organize and administer orientation, awareness or training programs for local education personnel on a similarly broad range of educational subjects, or can assist districts to plan and provide their own in-service training programs.

They can link a local district with other districts, either within New Jersey or anywhere else in the United States, whose experience and products can be helpful in meeting the needs of a local "consumer" district. This process is backed up and facilitated by New Jersey's diffusion and dissemination program which, along with all other states, validates such products or programs as being learner and cost effective. For many of these, consulting services, training services, and materials are available without cost directly from districts which produced them. There are hundreds of such nationally approved projects, many of which were developed by school districts within New Jersey.

The locations of EIC offices are:

EIC-South
Glassboro-Woodbury Road
P.O. Box 426
Pitman, NJ 08071

(609) 589-3410

EIC-Central
50 Lake Drive
Hightstown, NJ 08520

(609) 448-0484

EIC-Northwest
Halko Drive
Cedar Knolls, NJ 07920

(201) 539-0331

EIC-Northeast
86 South Harrison Street
East Orange, NJ 07208

(201) 677-9600

III. State Activity

The central State Department staff provides assistance by playing an important role in a national research and development diffusion network, which is coordinated by the National Institute of Education and the U.S. Office of Education. Other State procedures which may be helpful to districts in carrying out their T & E programs are results of Statewide Testing, conduct of Educational Audits, and support of educational research and development.

IV. Educational Associations

The contribution of the various educational associations in assisting school districts with a broad range of needs associated with T & E can also be of major significance. The associations have long been in the forefront of identifying and publishing information on a wide variety of educational subjects, and conducting training in these subjects as well.

Some selected educational associations which may provide helpful information are:

American Association for Health,
Physical Education and Recreation (AAHPER)
1201 16th Street, N.W.
Washington, DC 20036

American Council on the
Teaching of Foreign Languages (ACTFL)
62 Fifth Avenue
New York, NY 10011

American Educational Research Association (AERA)
1126 16th Street, N.W.
Washington, DC 20036

American Home Economics Association (AHEA)
2010 Massachusetts Avenue, N.W.
Washington, DC 20036

American Industrial Arts Association (AIAA)
1201 16th Street, N.W.
Washington, DC 20036

American Vocational Association (AVA)
1510 H Street N.W.
Washington, DC 20005

Educational Associations (continued)

Association for Supervision and
Curriculum Development (ASCD)
1201 16th Street, N.W.
Washington, DC 20036

International Reading Association (IRA)
Six Tyre Avenue
Newark, DE 19711

National Council for Geographic Education (NCGE)
115 North Marion Street
Oak Park, IL 60301

National Council on Measurement in Education (NCME)
Michigan State University
202 South Kedzie Hall
East Lansing, MI 48823

National Science Teachers Association (NSTA)
1201 16th Street, N.W.
Washington, DC 20036

National Council for the Social Studies (NCSS)
1201 16th Street, N.W.
Washington, DC 20036

National Council of Teachers of English (NCTE)
1111 Kenyon Road
Urbana, IL 61801

National Council of Teachers of Mathematics (NCTM)
1906 Association Drive
Reston, VA 22091

National Council of Technical Schools (NCTS)
1835 K Street, N.W.
Washington, DC 20006

National Education Association (NEA)
1201 16th Street, N.W.
Washington, DC 20036

New Jersey Education Association (NJEA)
180 West State Street
Trenton, NJ 08608

New Jersey School Boards Association (NJSBA)
P.O. Box 909
383 West State Street
Trenton, NJ 08605

Phi Delta Kappa
Eighth Street and Union Avenue
Bloomington, IN 47401

Teachers of English to Speakers
of Other Languages (TESOL)
School of Language and Linguistics
Georgetown University
Washington, DC 20007

You may, of course, know of other associations that would be helpful
resources for information on educational subjects.

V. Other possible sources of information and/or assistance include:

ERIC/TM
Educational Testing Service
Princeton, NJ 08540

Test Collection
Educational Testing Service
Princeton, NJ 08540